University of Namur

THESIS / THÈSE

MASTER EN SCIENCES MATHÉMATIQUES

Un algorithme de résolution de systèmes d'équations non lisses, basé sur la méthode de Newton

ORBAN, Frédérique

Award date:
1995

Awarding institution:
Universite de Namur

Facultés Universitaires Notre-Dame de la Paix
Namur
Faculté des Sciences

Un Algorithme de Résolution de Systèmes d'Equations non Lisses, basé sur la Méthode de Newton

Mémoire présenté pour l'obtention du grade
de Licencié en Sciences
mathématiques
par

Frédérique ORBAN

Promoteur: **J.-J. STRODIOT**

Année académique 1994-1995

Que Monsieur Jean-Jacques Strodiot trouve ici l'expression de ma gratitude pour son aimable collaboration et son aide.

Que tout ceux grâce à qui je suis arrivée là aujourd'hui en soient chaleureusement remerciés.

Je tiens en particulier à remercier ma famille, mes amis et Pascal pour leur soutien et leurs encouragements tout au long de mes études.

Un Algorithme de Résolution de Systèmes d'Equations non Lisses, basé sur la Méthode de Newton

ORBAN Frédérique

Résumé

L'objet de ce mémoire est la résolution de systèmes d'équations non linéaires définis par des fonctions localement lipschitziennes. Nous construisons un algorithme sur base de deux versions non lisses de la méthode de Newton classique, l'une convergeant globalement et l'autre localement de manière quadratique. Cet algorithme hybride conserve les propriétés de convergence de ses composantes. Les éléments qu'il requiert sont construits dans le cas du problème de complémentarité non linéaire.

Abstract

The purpose of this work is the resolution of systems of nonlinear equations defined by locally Lipschitzian functions. We construct an algorithm based on two nonsmooth versions of the classical Newton's method, the first one globally convergent, the second one, locally quadratically convergent. This hybrid algorithm keeps the convergence properties of its components. The elements it requires are built in the case of the nonlinear complementarity problem.

Mémoire de Licence en Sciences Mathématiques.
Juin 1995.
Unité d'Optimisation.
Promoteur: **J.-J. STRODIOT**

Errata

page 5 : Alors, il est clair que $\lambda^{(p)} \in [0,1]$ pour $p = 0, 1, \dots, m$ et $\sum_{p=0}^{m} \lambda^{(p)} = 1$

page 23 : Proposition 0.7 : --- où $F : \mathbb{R}^n \to \mathbb{R}^n$.

page 25 : --- le lagrangien augmenté d'un problème de programmation non linéaire de classe C^2 ---

page 31 : $\| x^k - x^* \| \leq \frac{\alpha}{1-\alpha} \| x^k - x^{k-1} \|$ (1.9).

page 33 : $\frac{\alpha - \alpha^{p+2}}{1-\alpha} \| x^k - x^{k-1} \| \leq \frac{\alpha}{1-\alpha} \| x^k - x^{k-1} \|$.

page 36 : remplacer les $\sum_{k=0}^{m}$ dans la preuve, par $\sum_{k=0}^{n}$ (proposition 1.3).

page 38 : --- les itérés x^k soient bien définis par (1.11) et (1.12), ---

page 42 : $V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix} \longrightarrow \begin{pmatrix} (\nabla f_i(\bar{x}))\left[r(\nabla f_i(\bar{x}))^T h + \alpha\right] + (\bar{\omega} + r f_i(\bar{x}))\, \nabla^2 f_i(\bar{x}) h \\ (\nabla f_i(\bar{x}))^T h \end{pmatrix}$

page 51 : $(2.18) \leq c\left[O(\| x^k - x^* \|^2) + O(\| x^k - x^* \|^2) \right]$.

page 72 : (4.5) : $\| x - x^* \| \leq 2c \| F(x) \|$.

TABLE DES MATIERES

INTRODUCTION

Le présent mémoire vise à résoudre le système d'équations non linéaires $F(x) = 0$, défini par une fonction $F: \mathbf{R}^n \to \mathbf{R}^n$ localement lipschitzienne. Parmi les nombreuses méthodes existant, celle de Newton est une des plus puissantes, lorsque la fonction F est continûment différentiable. Mais si F n'est que localement lipschitzienne, donc non nécessairement différentiable, la méthode de Newton n'est plus applicable et une modification s'impose. Nous en étudions différentes variantes, ainsi que leurs propriétés de convergence.

Dans notre analyse, nous supposons que la fonction F soit localement lipschitzienne. Les colonnes de sa matrice jacobienne sont constituées des gradients des composantes, pour autant qu'ils existent. La norme utilisée est la norme euclidienne.

Le chapitre préliminaire met en place une série de notions théoriques concernant les caractères semi-lisse et B-différentiable.

Le chapitre 1 présente une extension de la méthode de Newton classique au cas d'une fonction F non nécessairement différentiable: un élément du jacobien généralisé y remplace la dérivée de F. Cette extension comprend la version de la méthode de Newton basée sur la B-dérivée, en tant que cas spécial. Ensuite, la convergence locale de la méthode proposée est établie, à condition que F soit semi-lisse et que les matrices du jacobien généralisé soient inversibles. Enfin, il est démontré que la méthode étudiée peut être appliquée lors de la résolution de programmes non linéaires à l'aide du lagrangien augmenté.
Une amélioration de cette méthode est décrite au second chapitre. L'idée est de relâcher la condition imposée sur les éléments du jacobien généralisé. Il n'est en effet pas nécessaire que toutes ces matrices soient non singulières: nous l'exigerons uniquement pour un certain nombre d'entre elles. Le théorème de convergence locale (chapitre 1) est encore valable pour la méthode améliorée.

Ces deux chapitres traitent de méthodes possédant une bonne propriété de convergence locale: le taux en est superlinéaire, ou quadratique moyennant une hypothèse supplémentaire.

Le chapitre 3, quant à lui, généralise la méthode de Newton amortie afin d'obtenir un algorithme globalement convergent, applicable au cas non différentiable. Cette généralisation se fait en plusieurs étapes. Dans un premier temps, nous supposons que la fonction F soit différentiable dans toutes les directions. Un algorithme est présenté, ainsi qu'une étude de sa convergence. Mais la discontinuité potentielle de la dérivée directionnelle peut constituer un obstacle au bon déroulement de la méthode; une seconde étape consiste alors à établir un algorithme permettant de ne pas tenir compte de la dérivée citée. La convergence globale de l'algorithme généralisé est étudiée dans une dernière partie.

Le chapitre 4 construit un algorithme constitué de la méthode localement convergente vue au chapitre 2 et de l'algorithme généralisé établi au chapitre 3. Cet algorithme hybride conserve les propriétés de convergence des méthodes qui le composent: il converge dès lors globalement et, localement, de manière quadratique. Afin d'obtenir ces propriétés, il est utile de faire appel à un théorème d'attraction, valable dans des situations très générales. Ce théorème exprime le fait qu'un zéro de la fonction F peut être un tel zéro "fort" qu'il attire vers lui toute une suite d'itérés.

Enfin, l'objet du chapitre 5 est de déterminer une matrice et une fonction d'itération satisfaisant à certaines conditions, dans le cas particulier du problème de complémentarité non linéaire; l'existence de ces éléments est nécessaire au bon déroulement de l'algorithme hybride.

En conclusion, nous portons sur ce travail un regard critique et nous proposons quelques pistes de réflexion pour en élargir le champ d'application.

CHAPITRE PRELIMINAIRE

Notions de base

La première section de ce chapitre présentera une étude des fonctions de plusieurs variables, semi-lisses. Nous aborderons les concepts de jacobien généralisé et de fonctions semi-lisses et semi-lisses d'ordre p ($0 < p \leq 1$). Plusieurs moyens de caractériser des fonctions semi-lisses seront établis, ainsi que deux conditions suffisantes pour qu'une fonction vérifie cette propriété. Nous citerons pour terminer une condition nécessaire au caractère semi-lisse et une autre, nécessaire à la propriété "être semi-lisse d'ordre p".

Dans une seconde section, nous étudierons les fonctions B-différentiables et leurs liens avec la dérivée directionnelle et les fonctions semi-lisses. Les concepts de B-dérivée lipschitzienne, B-différentiabilité d'ordre 2 et dérivée directionnelle semi-continue ou semi-continue de degré 2 seront définis. Ensuite, nous verrons des propriétés obtenues sous l'hypothèse de B-différentiabilité. Nous terminerons par la notion de BD-régularité d'une fonction F et son utilité pour obtenir l'unicité locale d'une solution de l'équation $F(x) = 0$.

0.1 Fonctions de plusieurs variables, semi-lisses

Soit F: $\mathbf{R}^n \rightarrow \mathbf{R}^m$ une fonction localement lipschitzienne, c'est-à-dire

$$\forall\, x \in \mathbf{R}^n,\ \exists\, U_x, \text{ voisinage de } x,\ \exists\, L \geq 0 : \forall\, z, y \in U_x,$$
$$\| F(z) - F(y) \| \leq L \ \| z - y \|.$$

Dans ce chapitre, nous supposerons toujours que F soit une fonction localement lipschitzienne. D'après le théorème de Rademacher, F est alors différentiable presque partout.

Nous noterons D_F l'ensemble des points où F est différentiable et JF(x) la matrice jacobienne (n, m) usuelle des dérivées partielles, quand x est un point où les dérivées partielles nécessaires existent.

Définition 0.1

Le jacobien généralisé de F au point x (Clarke, [1]), noté ∂F(x), est donné par:

$$\partial F(x) = \text{co} \{ \lim_{\substack{x_i \rightarrow x \\ x_i \in D_F}} JF(x_i) \}. \tag{0.1}$$

Théorème 0.1 [théorème de la moyenne]

$\forall\, x, y \in \mathbf{R}^n$,

$$F(y) - F(x) \in \text{co}\, \delta F([x, y])\, (y - x), \tag{0.2}$$

où la partie de droite représente l'enveloppe convexe de tous les points de la forme $V(y - x)$, avec $V \in \delta F(u)$ pour un point $u \in [x, y]$.

□

Remarquons qu'il n'y a pas d'ambiguïté dans l'écriture de co $\delta F([x, y])\,(y - x)$, puisque $\{\text{co}\, \delta F([x, y])\}\,(y - x) = \text{co}\, \{\delta F([x, y])\,(y - x)\}$.

Proposition 0.1

Soient F: $\mathbf{R}^n \rightarrow \mathbf{R}^m$ une fonction localement lipschitzienne et x, h des vecteurs de $\mathbf{R}^n$.

Si $\lim_{\substack{V \in \delta F(x+th) \\ t \downarrow 0}} \{Vh\}$ existe, (0.3)

alors la dérivée directionnelle classique

$$F'(x;h) = \lim_{t\downarrow 0} \frac{F(x+th)-F(x)}{t}$$

existe et vaut

$$F'(x;h) = \lim_{\substack{V\in \delta F(x+th)\\ t\downarrow 0}} \{Vh\}. \tag{0.4}$$

PREUVE:

Soit $\{t_k\}$, $k \in \mathbf{N}$, une suite décroissante de réels positifs tendant vers zéro.

Puisque F est localement lipschitzienne, la suite

$$\left\{ \frac{F(x+t_k h)-F(x)}{t_k} \right\} \text{ est bornée.}$$

Notons a un point d'accumulation de cette suite. Alors,

$$a = \lim_{j\to\infty} \frac{F(x+t_j h)-F(x)}{t_j}$$

où $\{t_j\}$ est une sous-suite de $\{t_k\}$.

D'autre part, comme pour chaque j nous avons par (0.2) que

$$\frac{F(x+t_j h)-F(x)}{t_j} \in \text{co}\ \delta F([x, x+t_j h])\ h,$$

il suit du théorème de Carathéodory qu' il existe $t_j^{(p)} \in [0, t_j]$, $\lambda_j^{(p)} \in [0, 1]$ et $V_j^{(p)} \in \delta F([x, x+t_j^{(p)}h])$, pour $p = 0, 1,\ldots, m$, avec $\sum_{p=0}^{m} \lambda_j^{(p)} = 1$, tels que

$$\frac{F(x+t_j h)-F(x)}{t_j} = \sum_{p=0}^{m} \lambda_j^{(p)}\, V_j^{(p)}\, h\,.$$

En prenant si nécessaire une sous-suite, nous pouvons affirmer que pour tout $p = 0, 1,\ldots, m$, $\lambda_j^{(p)}$ converge vers un certain $\lambda^{(p)}$ lorsque j tend vers l'infini.

Alors, il est clair que $\lambda^{(k)} \in [0, 1]$ pour $k = 0, 1,\ldots, m$ et que $\sum_{k=0}^{m} \lambda^{(k)} = 1.$

Par conséquent,

$$a = \lim_{j\to\infty} \frac{F(x + t_j h) - F(x)}{t_j} = \lim_{j\to\infty} \sum_{p=0}^{m} \lambda_j^{(p)} V_j^{(p)} h$$

$$= \sum_{p=0}^{m} \lim_{j\to\infty} \lambda_j^{(p)} \lim_{j\to\infty} \{ V_j^{(p)} h \}, \qquad (0.5)$$

où $V_j^{(p)} \in \delta F([x, x+t_j^{(p)}h]),\ \forall\ p = 0, 1,.., m.$

Puisque la suite $\{t_j\}$ tend vers zéro et $t_j^{(p)} \in [0, t_j]$ pour tout $p = 0, 1,.., m$, nous obtenons

$$\lim_{j\to\infty} t_j^{(p)} = 0 \quad \text{et}$$

$$\lim_{\substack{V_j^{(p)} \in \delta F([x,\, x+t_j^{(p)}h]) \\ t_j\to\infty}} \{V_j^{(p)} h\} = \lim_{\substack{V\in \delta F(x+th) \\ t\downarrow 0}} \{Vh\}.$$

D'où l'expression (0.5) est encore égale à

$$\sum_{p=0}^{m} \lambda^{(p)} \lim_{\substack{V\in \delta F(x+th) \\ t\downarrow 0}} \{Vh\} = \lim_{\substack{V\in \delta F(x+th) \\ t\downarrow 0}} \{Vh\}.$$

Nous avons donc obtenu que tout point d'accumulation a de la suite

$$\left\{ \frac{F(x + t_k h) - F(x)}{t_k} \right\}$$

est tel que

$$a = \lim_{\substack{V\in \delta F(x+th) \\ t\downarrow 0}} \{Vh\}.$$

Nous en déduisons que la suite entière converge vers cette limite lorsque k tend vers l'infini. Par conséquent, la dérivée directionnelle classique F'(x; h) existe et l'égalité (0.4) est vérifiée. □

Définition 0.2

On dit que F: $\mathbf{R}^n \to \mathbf{R}^m$ est <u>semi-lisse en x</u> si F est localement lipschitzienne en x et $\lim_{\substack{V \in \delta F(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\}$ existe pour tout h dans $\mathbf{R}^n$.

Remarque: l'existence de cette limite est une affirmation plus forte que (0.3). Elle implique que la dérivée directionnelle d'Hadamard vérifie

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{F(x + th') - F(x)}{t} = \lim_{\substack{V \in \delta F(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\}. \tag{0.6}$$

Les fonctions convexes et les fonctions lisses sont des exemples de fonctions semi-lisses; le produit scalaire, la somme et les fonctions composées de fonctions semi-lisses sont encore des fonctions semi-lisses.

Proposition 0.2

Soit F: $\mathbf{R}^n \to \mathbf{R}^m$ une fonction localement lipschitzienne. Si F'(x; h) existe en x pour tout h, alors:

(i) F'(x; •) est lipschitzienne;

(ii) $\forall$ h, $\exists$ V $\in$ $\delta F(x)$ tel que F'(x; h) = Vh.

PREUVE:

Soit L la constante de Lipschitz de F au voisinage de x.

Montrons d'abord que F'(x; •) est lipschitzienne. Pour tout h, h' dans $\mathbf{R}^n$, nous avons successivement

$$\| F'(x; h) - F'(x; h') \|$$

$$= \left\| \lim_{t \downarrow 0} \frac{F(x + th) - F(x + th')}{t} \right\| \qquad \text{(hypothèse d'existence)}$$

$$= \lim_{t \downarrow 0} \frac{\| F(x + th) - F(x + th') \|}{t}$$

$$\leq \lim_{t \downarrow 0} L \frac{\| t(h - h') \|}{t} \qquad \text{(F localement lipschitzienne)}$$

$$= L \| h - h' \|.$$

La première assertion de la proposition est ainsi prouvée.

D'autre part, par définition de la dérivée directionnelle nous avons

$$F'(x; h) \; = \; \lim_{t \downarrow 0} \frac{F(x + th) - F(x)}{t} \, .$$

Il existe donc une suite $\{t_k\}$, $k \in \mathbf{N}$, de nombres positifs décroissant vers zéro, telle que

$$F'(x; h) \; = \; \lim_{k \to \infty} \frac{F(x + t_k h) - F(x)}{t_k} \, .$$

Par (0.2), $F(x + t_k h) - F(x) \in \text{co}\; \delta F([x, x+t_k h])\,(t_k h)$, ce qui implique que, pour tout k, il existe une matrice V_k dans $\text{co}\; \delta F([x, x+t_k h])$, telle que

$$F(x + t_k h) - F(x) \; = \; V_k \, t_k \, h.$$

On en déduit qu'il existe une suite décroissante $\{t_k\}$ tendant vers zéro et une suite $\{V_k\}$, $V_k \in \text{co}\; \delta F([x, x+t_k h])$ pour tout k, telles que

$$F'(x; h) \; = \; \lim_{k \to \infty} \{V_k h\} \, .$$

Puisque F est localement lipschitzienne, la suite $\{V_k\}$ est bornée. Comme h est arbitraire, l'égalité précédente permet d'affirmer que toutes les sous-suites de $\{V_k\}$ convergent vers la même limite V et donc que la suite entière converge vers V.
Le jacobien généralisé $\delta F(x)$ étant fermé il contient V. En conclusion, nous avons trouvé une matrice V dans $\delta F(x)$ telle que $F'(x; h) = Vh$, ce qui achève la preuve. □

Le théorème suivant donne différentes caractérisations de fonctions semi-lisses.

<u>Théorème 0.2</u> [Caractérisation de fonctions semi-lisses]

Soit $F\colon \mathbf{R}^n \to \mathbf{R}^m$ une fonction localement lipschitzienne. Les propositions suivantes sont équivalentes:

(i) F est semi-lisse en x;

(ii) la limite (voir (0.6)) $\lim_{\substack{V \in \delta F(x+th') \\ h' \to h,\, t \downarrow 0}} \{Vh'\}$ est uniforme pour tout h tel que $\| h \| = 1$;

(iii) la limite (voir (0.4)) $\lim_{\substack{V\in \delta F(x+th) \\ t\downarrow 0}} \{Vh\}$ est uniforme pour tout h

tel que $\| h \| = 1$;

(iv) $\forall\, h \to 0,\ \forall\, V \in \delta F(x+h),\ Vh - F'(x; h) = o(\| h \|)$ (0.7)

(v) $\lim_{\substack{x+h\in D_F \\ h\to 0}} \dfrac{F'(x+h; h) - F'(x; h)}{\| h \|} = 0.$ (0.8)

PREUVE:

$\boxed{(i) \Rightarrow (ii)}$

Comme le point de convergence en (ii) ne peut être que F'(x; h) (voir proposition 0.1), il suffit de prouver que:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 : \forall\, h \text{ tel que } \| h \| = 1, \forall\, h', \forall\, t, \forall\, V \in \delta F(x+th'),$$

$$\| h - h' \| < \delta \text{ et } 0 < t < \delta \Rightarrow \| Vh' - F'(x; h) \| < 2\varepsilon. \qquad (0.9)$$

Supposons par l'absurde que (0.9) ne soit pas vérifiée, c'est-à-dire que:

$$\exists\, \varepsilon > 0 : \forall\, \delta > 0, \exists\, h \text{ tel que } \| h \| = 1, \exists\, h', \exists\, t, \exists\, V \in \delta F(x+th') \text{ tels que}$$

$$\| h - h' \| < \delta,\ 0 < t < \delta \text{ et } \| Vh' - F'(x; h) \| \geq 2\varepsilon.$$

Choisissant $\delta = \frac{1}{k}$, $k \in \mathbf{N}$, nous obtenons:

$\exists\, \{h^k\}$, $k \in \mathbf{N}$ telle que $\| h^k \| = 1\ \forall\, k$, $\exists\, \{h'^k\}$, $\exists\, \{t_k\}$, $k \in \mathbf{N}$, $\exists\, \{V_k\}$,

$k \in \mathbf{N}$, telle que $V_k \in \delta F(x+t_k h'^k)$ pour tout k, de telle façon que:

$\| h^k - h'^k \| \to 0$, $t_k \downarrow 0$ lorsque $k \to \infty$

et $\| V_k h'^k - F'(x; h^k) \| \geq 2\varepsilon.$ (0.10)

En passant si nécessaire à une sous-suite, nous pouvons affirmer que la suite $\{h^k\}$ converge vers une certaine limite h.

Comme

$$0 \leq \| h'^k - h \| \leq \| h'^k - h + h^k - h^k \| \leq \| h^k - h'^k \| + \| h^k - h \|,$$

la suite $\{h'^k\}$ converge aussi vers h.

Le point (i) de la proposition 0.2 nous permet de déduire de (0.10) que pour tout k assez grand,

$$\| V_k h'^k - F'(x; h) \| \geq \varepsilon. \tag{0.11}$$

En effet, nous avons successivement:

$$
\begin{aligned}
2\,\varepsilon &\leq \| V_k h'^k - F'(x; h^k) \| && \text{(expression (0.10))} \\
&= \| V_k h'^k - F'(x; h) + F'(x; h) - F'(x; h^k) \| \\
&\leq \| V_k h'^k - F'(x; h) \| + \| F'(x; h) - F'(x; h^k) \| \\
&\leq \| V_k h'^k - F'(x; h) \| + \varepsilon,
\end{aligned}
$$

où la dernière inégalité est valable pour tout k suffisamment grand, puisque $F'(x; \bullet)$ est lipschitzienne en vue de la proposition 0.2.

Enfin, (0.11) contredit l'hypothèse (i), c'est-à-dire le fait que

$$\lim_{\substack{V \in \delta F(x+th') \\ h' \to h,\ t \downarrow 0}} \{Vh'\} \text{ existe pour tout h dans } \mathbf{R}^n .$$

D'où l'assertion (0.9) est vérifiée et donc également (ii).

$\boxed{\text{(ii)} \Rightarrow \text{(iii)}}$

Nous savons que l'hypothèse est équivalente à (0.9).
De plus, il est évident que (0.9) implique:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 : \forall \text{ h tel que } \| h \| = 1, \forall\, t, \forall\, V \in \delta F(x+th),$$
$$0 < t < \delta \Rightarrow \| Vh - F'(x; h) \| < 2\varepsilon,$$

c'est-à-dire l'assertion (iii).

$\boxed{\text{(iii)} \Rightarrow \text{(iv)}}$

Nous devons prouver que

$$\lim_{\substack{V \in \delta F(x+h') \\ h' \to 0}} \frac{Vh' - F'(x; h')}{\| h' \|} = 0,$$

ou encore

$$\forall \varepsilon > 0, \exists \delta > 0 : \forall h' \text{ tel que } 0 < \| h' \| < \delta, \forall V \in \delta F(x+h'),$$

$$\frac{\| Vh' - F'(x; h') \|}{\| h' \|} < \varepsilon. \qquad (0.12)$$

Par hypothèse, $\lim_{\substack{V \in \delta F(x+th) \\ t \downarrow 0}} \{Vh\} = F'(x; h)$ pour tout h tel que $\| h \| = 1$.

Ceci peut s'écrire:

$$\forall \varepsilon > 0, \exists \delta > 0 : \forall t \text{ tel que } 0 < t < \delta, \forall h \text{ tel que } \| h \| = 1,$$

$$\forall V \in \delta F(x+th), \ \| Vh - F'(x; h) \| < \varepsilon. \qquad (0.13)$$

Par conséquent, en prenant $t = \| h' \|$ et $h = \dfrac{h'}{\| h' \|}$ dans (0.13), l'homogénéité de la dérivée directionnelle nous permet d'obtenir:

$$\forall \varepsilon > 0, \exists \delta > 0 : \forall h' \text{ tel que } 0 < \| h' \| < \delta, \forall V \in \delta F(x+h'),$$

$$\frac{\| Vh' - F'(x; h') \|}{\| h' \|} = \| Vh - F'(x; h) \| < \varepsilon.$$

L'expression (0.12) est ainsi prouvée.

$\boxed{\text{(iv)} \Rightarrow \text{(i)}}$

Supposons par l'absurde que F ne soit pas semi-lisse en x. Alors, il existe un vecteur h de $\mathbf{R}^n$, des suites $\{h^k\}$ convergeant vers h, $\{t_k\}$ décroissant vers zéro et $\{V_k\}$ vérifiant $V_k \in \delta F(x+t_k h^k)$ pour tout k, ainsi qu'un réel ε strictement positif, tels que

$$\| V_k h^k - F'(x; h) \| \geq 2\varepsilon \qquad \forall k. \qquad (0.14)$$

D'une manière analogue à celle utilisée dans la première partie de cette démonstration pour passer de (0.10) à (0.11), la proposition 0.2 nous permet de déduire de (0.14) que pour tout k assez grand,

$$\| V_k h^k - F'(x; h^k) \| \geq \varepsilon.$$

Nous obtenons ainsi une contradiction avec l'hypothèse (0.7) qui entraîne que l'expression

$$V_k\, h^k - F'(x; h^k)$$

converge vers zéro lorsque k tend vers l'infini.

$\boxed{(iv) \Rightarrow (v)}$

Cette implication est immédiate grâce au point (ii) de la proposition 0.2.
En effet, cette proposition affirme que pour tout h, il existe une matrice V dans $\delta F(x+h)$ telle que $F'(x + h; h) = Vh$.
Or, tout V dans $\delta F(x+h)$ vérifie par hypothèse:

$$\lim_{\substack{V \in \delta F(x+h) \\ h \to 0}} \frac{Vh - F'(x; h)}{\| h \|} = 0,$$

d'où

$$\lim_{\substack{x+h \in D_F \\ h \to 0}} \frac{F'(x + h; h) - F'(x; h)}{\| h \|} = \lim_{\substack{V \in \delta F(x+h) \\ h \to 0}} \frac{Vh - F'(x; h)}{\| h \|} = 0.$$

$\boxed{(v) \Rightarrow (iv)}$

Notre hypothèse peut se réécrire:

$$\forall\, \varepsilon > 0, \exists\, \delta > 0 \text{ tel que } \forall\, h \in \mathbf{R}^n,$$
$$\| h \| < \delta \text{ et } x + h \in D_F \Rightarrow \| F'(x + h; h) - F'(x; h) \| < \varepsilon \| h \|. \qquad (0.15)$$

Prenons h et V tels que $\| h \| \leq \frac{1}{2}\delta$ et $V \in \delta F(x+h)$; afin de prouver l'assertion (iv), nous allons montrer que

$$\| Vh - F'(x; h) \| \leq 5\, \varepsilon \| h \|. \qquad (0.16)$$

Par définition du jacobien généralisé $\delta F(x)$, nous avons que

$$Vh \in \mathrm{co}\,\{\lim_{\substack{h^i \to h \\ x+h^i \in D_F}} F'(x + h^i; h)\}$$

et par le théorème de Carathéodory, nous obtenons:

$$\exists\, \lambda_0, \lambda_1,.., \lambda_m \geq 0, \ \sum_{k=0}^{m} \lambda_k = 1 \text{ tels que } Vh = \sum_{k=0}^{m} \lambda_k X^{(k)}$$

où $X^{(k)} \in \{\lim_{\substack{h^i \to h \\ x+h^i \in D_F}} F'(x + h^i; h)\} \ \ \forall\, k = 0, 1,.., m,$

c'est-à-dire:

$$\forall\, k = 0, 1,.., m, \exists\, \lambda_k \geq 0, \ \sum_{k=0}^{m} \lambda_k = 1 \text{ tels que } \forall\, \varepsilon > 0, \exists\, \eta_k > 0,$$

$$\exists\, h^{(k)}, \ h^{(k)} \in \{h^i\}, \| h^{(k)} - h \| < \eta_k \text{ et } x + h^{(k)} \in D_F,$$

vérifiant en particulier:

$$\| h^{(k)} - h \| \leq \min \{ \tfrac{1}{2}\delta, \| h \|, \varepsilon\frac{\| h \|}{L} \}, \tag{0.17}$$

$$\text{tels que } Vh = \sum_{k=0}^{m} \lambda_k X^{(k)} \text{ et } \| X^{(k)} - F'(x + h^{(k)}; h) \| < \varepsilon \| h \|,$$

où L est la constante de Lipschitz de F au voisinage de x.

Nous en déduisons:

$$\exists\, \lambda_0, \lambda_1,.., \lambda_m \geq 0, \ \sum_{k=0}^{m} \lambda_k = 1, \exists\, h^{(0)}, h^{(1)},.., h^{(m)} \text{ tels que}$$

$$\| h^{(k)} - h \| \leq \min \{\tfrac{1}{2}\delta, \| h \|, \varepsilon\frac{\| h \|}{L} \}, x + h^{(k)} \in D_F \text{ pour } k = 0, 1,.., m$$

$$\text{et } \| Vh - \sum_{k=0}^{m} \lambda_k F'(x + h^{(k)}; h) \| < \varepsilon \| h \|. \tag{0.18}$$

Par la proposition 0.2, (i) et en utilisant l'inégalité (0.15) avec $h^{(k)}$, nous obtenons successivement:

$$\| \sum_{k=0}^{m} \lambda_k F'(x + h^{(k)}; h) - \sum_{k=0}^{m} \lambda_k F'(x; h) \|$$

$$\leq \sum_{k=0}^{m} \lambda_k \Big[\| F'(x + h^{(k)}; h^{(k)}) - F'(x + h^{(k)}; h) \|$$

$$+ \| F'(x + h^{(k)}; h^{(k)}) - F'(x; h^{(k)}) \| + \| F'(x; h) - F'(x; h^{(k)}) \| \Big]$$

$$\leq \sum_{k=0}^{m} \lambda_k \Big[L \| h^{(k)} - h \| + \varepsilon \| h^{(k)} \| + L \| h^{(k)} - h \| \Big],$$

où la condition supplémentaire (0.17) imposée sur les $h^{(k)}$ a permis d'utiliser la relation (0.15) pour $h^{(k)}$, car $\| h^{(k)} \| \leq \| h^{(k)} - h \| + \| h \| \leq \frac{1}{2} \delta + \frac{1}{2} \delta = \delta$,

$$\leq \sum_{k=0}^{m} \lambda_k \Big[\varepsilon \| h \| + 2\varepsilon \| h \| + \varepsilon \| h \| \Big],$$

car par (0.17), $L \| h^{(k)} - h \| \leq L \frac{\varepsilon \| h \|}{L} = \varepsilon \| h \|$
et $\varepsilon \| h^{(k)} \| \leq \varepsilon \| h^{(k)} - h \| + \varepsilon \| h \| \leq 2\varepsilon \| h \|$,

$$= 4\varepsilon \| h \|. \qquad (0.19)$$

Finalement, (0.16) se déduit de (0.18) et (0.19):

$$\| Vh - F'(x; h) \|$$

$$\leq \| Vh - \sum_{k=0}^{m} \lambda_k F'(x + h^{(k)}; h) \| + \| \sum_{k=0}^{m} \lambda_k F'(x + h^{(k)}; h) - F'(x; h) \|$$

$$\leq \varepsilon \| h \| + 4\varepsilon \| h \| = 5\varepsilon \| h \|.$$ □

Corollaire 0.3

Soit F: $\mathbf{R}^n \to \mathbf{R}^m$ une fonction localement lipschitzienne.
Si chaque composante de F est semi-lisse en x,
alors F est semi-lisse en x.

PREUVE:
Prenons $x + h$ dans D_F et h convergeant vers zéro. Nous allons montrer que l'égalité (0.8) du théorème 0.3 est vérifiée, c'est-à-dire que:

$$\lim_{\substack{x+h \in D_F \\ h \to 0}} \frac{F'(x + h; h) - F'(x; h)}{\| h \|} = 0.$$

Par définition de la norme euclidienne, nous savons que:

$$\| F'(x + h; h) - F'(x; h) \| \leq \sum_{i=1}^{m} | F_i'(x + h; h) - F_i'(x; h) |.$$

En appliquant (0.8) à chaque F_i, nous obtenons:

$$\| F'(x + h; h) - F'(x; h) \|$$

$$\leq \sum_{i=1}^{m} | F_i'(x + h; h) - F_i'(x; h) | = o (\| h \|).$$

L'expression (0.8) est ainsi vérifiée par F et la thèse est prouvée. □

Définition 0.3
La dérivée de Fréchet en x, $F'(x)$, est dite forte si

$$\lim_{\substack{y \to x \\ z \to x}} \frac{F(x) - F(y) - F'(x)(z - y)}{\| z - y \|} = 0. \qquad (0.20)$$

Corollaire 0.4

Si F a une dérivée forte de Fréchet en x,
alors F est semi-lisse en x.

□

Remarques
(1) Si pour tout V dans $\delta F(x+h)$ et h tendant vers zéro, $Vh - F'(x; h) = O (\| h \|^{1+p})$, où $0 < p \leq 1$, alors on dit que F est semi-lisse d'ordre p en x.

Notons que la propriété "être semi-lisse d'ordre p " ($0 < p \leq 1$) entraîne la propriété "être semi-lisse".

(2) Par le théorème de la moyenne (théorème 0.1), si F est semi-lisse en x, alors $\forall\, h \to 0$,

$$F(x + h) - F(x) - F'(x; h) = o(\| h \|); \tag{0.21}$$

si F est semi-lisse d'ordre p en x, alors $\forall\, h \to 0$,

$$F(x + h) - F(x) - F'(x; h) = O(\| h \|^{1+p}).$$

0.2 Fonctions B-différentiables

Définition 0.4

Une fonction F: $\mathbf{R}^n \to \mathbf{R}^m$ est dite B-différentiable (B pour Bouligand) en un point x si il existe une fonction BF(x): $\mathbf{R}^n \to \mathbf{R}^n$ positivement homogène de degré 1 (c'est-à-dire $BF(x)(th) = t\, BF(x)h$ pour tout h dans $\mathbf{R}^n$ et tout t positif), appelée B-dérivée de F en x, telle que

$$\lim_{h\to 0} \frac{F(x + h) - F(x) - BF(x)h}{\| h \|} = 0. \tag{0.22}$$

Remarques

Dans un espace euclidien de dimension finie, $\mathbf{R}^n$, une fonction localement lipschitzienne F est B-différentiable en x si et seulement si elle est différentiable dans toutes les directions en x; dans ce cas, la dérivée directionnelle et la B-dérivée sont identiques (voir proposition 1 des annexes). Par conséquent, nous ne ferons pas de distinction entre ces deux dérivées et l'égalité (0.22) est équivalente à

$$\lim_{h\to 0} \frac{F(x + h) - F(x) - F'(x; h)}{\| h \|} = 0.$$

Proposition 0.3

Si F est une fonction semi-lisse en x, alors F est B-différentiable en x.

PREUVE:

Puisque F est semi-lisse, nous savons par (0.21) que

$$F(x + h) - F(x) - F'(x; h) = o(\| h \|)$$

pour tout h tendant vers zéro, c'est-à-dire

$$\lim_{h \to 0} \frac{F(x + h) - F(x) - F'(x; h)}{\| h \|} = 0.$$

Pour obtenir le résultat désiré, il suffit alors de remarquer que cette dernière égalité est équivalente à l'expression (0.22). □

Définition 0.5

La B-dérivée $BF(\bullet)$ est dite lipschitzienne en x si il existe une voisinage U_x de x et une constante positive L tels que pour tout y dans U_x et tout h de norme 1 dans $\mathbf{R}^n$,

$$\| (BF(y) - BF(x)) h \| \leq L \| y - x \|. \tag{0.23}$$

Notons que si $BF(\bullet)$ est lipschitzienne en x, alors F est localement lipschitzienne en x.

Définition 0.6

Soit $F: \mathbf{R}^n \to \mathbf{R}^m$ une fonction B-différentiable en un point x. F est B-différentiable de degré 2 en x si $F(x + h) = F(x) + F'(x; h) + O(\| h \|^2)$. (0.24)

Définition 0.7

Supposons que $F: \mathbf{R}^n \to \mathbf{R}^m$ soit B-différentiable dans un voisinage de x. La dérivée directionnelle $F'(\bullet; \bullet)$ est semi-continue en x si pour tout ε strictement positif, il existe un voisinage U de x tel que pour tout vecteur (x + h) de U,

$$\| F'(x + h; h) - F'(x; h) \| \leq \varepsilon \| h \|. \tag{0.25}$$

La dérivée directionnelle $F'(\bullet; \bullet)$ est semi-continue de degré 2 en x s'il existe une constante L et un voisinage U de x tels que pour tout vecteur (x + h) de U,

$$\| F'(x + h; h) - F'(x; h) \| \leq L \| h \|^2. \tag{0.26}$$

Proposition 0.4

Soit $F: \mathbf{R}^n \to \mathbf{R}^m$ une fonction localement lipschitzienne.
Supposons que F soit différentiable dans toutes les directions au voisinage de x.

Alors les assertions suivantes sont équivalentes:

(i) F est semi-lisse en x;

(ii) $F'(\bullet; \bullet)$ est semi-continue en x.

PREUVE:

$\boxed{(i) \Rightarrow (ii)}$

Puisque F est localement lipschitzienne et différentiable dans toutes les directions, nous obtenons par la proposition 0.2 qu'il existe une matrice V' dans $\delta F(x+h)$, telle que $F'(x + h; h) = V'h$.

Par le théorème 0.2, l'hypothèse est équivalente à l'affirmation suivante:

$$\forall\, V \in \delta F(x+h),\ \forall\, h \rightarrow 0,\ Vh - F'(x; h) = o(\| h \|).$$

Donc, V' vérifie

$$V'h - F'(x; h) = o(\| h \|),$$

d'où

$$F'(x + h; h) - F'(x; h) = o(\| h \|).$$

L'assertion (ii) est ainsi prouvée.

$\boxed{(ii) \Rightarrow (i)}$

Nous savons grâce au théorème 0.2 que la thèse (i) équivaut à la relation suivante:

$$\lim_{\substack{h \rightarrow 0 \\ x+h \in D_F}} \frac{F'(x + h; h) - F'(x; h)}{\| h \|} = 0.$$

Cette dernière égalité est obtenue à l'aide de (0.25); par conséquent, F est bien semi-lisse en x. □

Proposition 0.5

Soit F: $\mathbf{R}^n \rightarrow \mathbf{R}^m$ une fonction localement lipschitzienne.

Supposons que F soit différentiable dans toutes les directions au voisinage de x.

Les propositions suivantes sont alors équivalentes:

(i) $F'(\bullet; \bullet)$ est semi-continue de degré 2 en x;

(ii) Pour tout V dans $\delta F(x+h)$, pour tout $h \to 0$, $Vh - F'(x; h) = O(\| h \|^2)$,

c'est-à-dire F est semi-lisse d'ordre 1 en x. (0.27)

Si l'une de ces deux assertions est vérifiée, alors F est B-différentiable de degré 2 en x.

<u>PREUVE</u>:

(ii) ⇒ (i)

D'après la proposition 0.2, il existe une matrice V' dans $\delta F(x+h)$ telle que $F'(x + h; h) = V'h$.

Par hypothèse, il existe un voisinage U de x et une constante L de sorte que, pour tout vecteur x + h de U, pour tout V appartenant à $\delta F(x+h)$,

$$\| Vh - F'(x; h) \| \leq L \| h \|^2. \tag{0.28}$$

En particulier, l'inégalité (0.28) peut être appliquée à V'. Nous obtenons alors qu'il existe un voisinage U de x et une constante L, tels que pour tout x + h dans U,

$$\begin{aligned} \| F'(x + h; h) - F'(x; h) \| &= \| V'h - F'(x; h) \| \\ &\leq L \| h \|^2. \end{aligned}$$

La dérivée directionnelle $F'(\bullet, \bullet)$ est donc semi-continue de degré 2 en x.

(i) ⇒ (ii)

Par hypothèse, il existe une constante L_1 et un voisinage U de x tels que pour tout vecteur x + h de U, $\| F'(x + h; h) - F'(x; h) \| \leq L_1 \| h \|^2$, c'est-à-dire qu'il existe un réel δ_1 strictement compris entre 0 et 1 et une constante L_1, de sorte que pour tout vecteur h de $\mathbf{R}^n$ vérifiant $\| h \| < \delta_1$,

$$\| F'(x + h; h) - F'(x; h) \| \leq L_1 \| h \|^2. \tag{0.29}$$

De plus, puisque $F'(x; \bullet)$ est lipschitzienne (voir proposition 0.2), nous avons aussi qu'il existe un réel δ_2 strictement compris entre 0 et 1 et une constante L_2, tels que pour tout h, h', h" dans $\mathbf{R}^n$ avec $\| h \| < \delta_2$,

$$\| F'(x + h; h') - F'(x + h; h'') \| \leq L_2 \| h' - h'' \|. \tag{0.30}$$

Nous déduisons de (0.29) et (0.30) qu'il existe un réel δ dans l'intervalle ouvert (0, 1) et une constante L, tels que pour tout h, h' et h'' de $\mathbf{R}^n$, avec $\| h \| < 2\delta$,

$$\| F'(x + h; h) - F'(x; h) \| \leq L \| h \|^2 \qquad (0.31)$$

et

$$\| F'(x + h; h') - F'(x + h; h'') \| \leq L \| h' - h'' \|. \qquad (0.32)$$

Prenons maintenant un vecteur h de $\mathbf{R}^n$ tel que $\| h \| < \delta$ et une matrice V dans $\delta F(x+h)$.
Nous allons montrer que $Vh - F'(x; h) = O(\| h \|^2)$.
D'après la définition du jacobien généralisé et le théorème de Carathéodory, nous obtenons:

$\exists\, h^o, h^1, .., h^m$ tels que $\forall\, p = 0, 1, .., m$, F soit différentiable en $x + h^p$,

$$\| h^p - h \| \leq \min \{\delta, \varepsilon \| h \|^2\} \qquad (0.33)$$

$$\text{et } \| Vh - \sum_{p=0}^{m} \lambda_p JF(x + h^p)\, h \| \leq \varepsilon L \| h \|^2, \qquad (0.34)$$

$$\text{où } \lambda_p \geq 0\ \forall\, p, \ \sum_{p=0}^{m} \lambda_p = 1 \text{ et } 0 < \varepsilon < 1.$$

Nous avons alors successivement, pour tout p = 0, 1,.., m:

$$\| JF(x + h^p)\, h^p - F'(x; h^p) \|$$

$\leq L \| h^p \|^2$, où la condition (0.33) a permis d'utiliser (0.31), car $\| h^p \| \leq \| h^p - h \| + \| h \| \leq 2\delta$,

$\leq L (\| h \| + \| h^p - h \|)^2$

$\leq L (\| h \| + \varepsilon \| h \|^2)^2$ par (0.33),

$\leq L (\| h \| + \varepsilon \| h \|)^2$ car $\| h \| < \delta < 1$,

$\leq L (1 + 3\varepsilon) \| h \|^2$ puisque $\varepsilon < 1$, d'où $\varepsilon^2 < \varepsilon$. (0.35)

Ensuite,

$$\| JF(x + h^p)\, h^p - JF(x + h^p)\, h \|$$

$\leq L \| h^p - h \|$ par (0.32),

$$\leq \varepsilon L \| h \|^2 \qquad \text{par (0.33).} \qquad (0.36)$$

Ajoutons encore que

$$\| F'(x; h^p) - F'(x; h) \|$$

$$\leq L \| h^p - h \| \leq \varepsilon L \| h \|^2 \qquad \text{grâce à (0.32) et (0.33).} \qquad (0.37)$$

Enfin, nous pouvons conclure:

$$\| Vh - F'(x; h) \|$$

$$\leq \| Vh - \sum_{p=0}^{m} \lambda_p JF(x + h^p) h \| + \sum_{p=0}^{m} \lambda_p \| JF(x + h^p) h - F'(x; h) \|$$

$$\leq \varepsilon L \| h \|^2 + \sum_{p=0}^{m} \lambda_p \| JF(x + h^p) h - F'(x; h) \| \qquad \text{par (0.34),}$$

$$\leq \varepsilon L \| h \|^2 + \sum_{p=0}^{m} \lambda_p \Big[\| JF(x + h^p) h - JF(x + h^p) h^p \|$$

$$+ \| JF(x + h^p) h^p - F'(x; h^p) \| + \| F'(x; h^p) - F'(x; h) \| \Big]$$

$$\leq \varepsilon L \| h \|^2 + \sum_{p=0}^{m} \lambda_p \left[\varepsilon L \| h \|^2 + L (1 + 3\varepsilon) \| h \|^2 + \varepsilon L \| h \|^2 \right],$$

grâce aux résultats (0.36), (0.35) et (0.37),

$$= \varepsilon L \| h \|^2 + L (1 + 5\varepsilon) \| h \|^2$$

$$= (1 + 6\varepsilon) L \| h \|^2.$$

Par conséquent, nous avons bien prouvé que pour tout V dans $\delta F(x+h)$, pour h tendant vers zéro,

$$Vh - F'(x; h) = O(\| h \|^2).$$

Pour terminer, supposons que l'assertion (i) soit vérifiée. Définissons $\phi(t) \equiv F(x + th)$. Alors, ϕ est localement lipschitzienne dans [0, 1]. Chaque composante ϕ_i est donc absolument continue, d'où la fonction ϕ est différentiable presque partout dans [0, 1] et

$$F(x + h) - F(x) = \phi(1) - \phi(0) = \int_0^1 \phi'(t)\, dt$$

$$= \int_0^1 F'(x + th; h)\, dt$$

$$= \int_0^1 [\, F'(x; h) + O(\| th \|^2)]\, dt \qquad \text{par hypothèse,}$$

$$= F'(x; h) + O(\| h \|^2).$$

En résumé, nous avons obtenu

$$F(x + h) = F(x) + F'(x; h) + O(\| h \|^2).$$

F est donc B-différentiable de degré 2 en x.
Ceci achève la preuve de la proposition. □

Définition 0.8

Soit F une fonction différentiable dans toutes les directions en x. F est BD-régulière en x (BD pour B-dérivée) si pour tout h non nul dans $\mathbf{R}^n$, la dérivée directionnelle $F'(x; h)$ est différente de zéro.

Proposition 0.6

Si $F: \mathbf{R}^n \to \mathbf{R}^m$ est une fonction BD-régulière en x,
alors il existe une constante c telle que pour tout vecteur h de $\mathbf{R}^n$, $\| h \| \leq c \, \| F'(x; h) \|$. (0.38)

PREUVE:

Supposons par l'absurde que la thèse ne soit pas vérifiée. Nous pouvons alors affirmer que pour tout c, il existe un vecteur h de $\mathbf{R}^n$ tel que

$$\| h \| > c \, \| F'(x; h) \|,$$

c'est-à-dire

$$1 > c \, \| F'(x; \frac{h}{\| h \|}) \|.$$

Choisissons c = k pour k = 1, 2,.. Nous obtenons alors qu'il existe un suite $\{h^k\}$ dans $\mathbf{R}^n$, avec $\| h^k \| = 1$ pour tout k, telle que nous ayons

$$\frac{1}{k} > \| F'(x; h^k) \|.$$

Puisque cette suite $\{h^k\}$ est bornée, elle admet un point d'accumulation h, de norme 1. En prenant si nécessaire une sous-suite, nous pouvons prétendre que $\{h^k\}$ converge vers h. Ainsi, lorsque k tend vers l'infini, nous déduisons de l'inégalité ci-dessus que $F'(x; h) = 0$.

Nous avons donc trouvé un vecteur h non nul de $\mathbf{R}^n$ tel que $F'(x; h) = 0$. Ceci contredit l'hypothèse; la thèse est donc vérifiée. □

Remarque

Il est facile de voir que la réciproque de cette proposition est vraie.

En effet, supposons que la condition (0.38) soit satisfaite. Notons d'une part que la constante c est différente de zéro, car si c valait zéro, la relation (0.38) ne serait pas vérifiée pour un vecteur h de norme un. Or, cette relation doit être valable pour tout h. D'autre part, il est évident que F est BD-régulière en x, c'est-à-dire que pour tout h non nul dans $\mathbf{R}^n$, la dérivée $F'(x; h)$ est différente de zéro, car sinon nous obtiendrions l'inégalité $\| h \| \leq 0$ pour $h \neq 0$.

La proposition suivante ne sera pas directement utilisée dans la suite, mais elle explique certaines relations entre la BD-régularité et des solutions de l'équation $F(x) = 0$: elle assure l'unicité locale d'une solution x*, à condition que F soit semi-lisse et BD-régulière en x*.

Proposition 0.7

Soit x* une solution de $F(x) = 0$, où $F: \mathbf{R}^n \to \mathbf{R}^m$ est une fonction localement lipschitzienne.
Si F est semi-lisse et BD-régulière en x*, alors il existe un voisinage de x* tel que x* soit solution unique dans ce voisinage.

PREUVE:

Si la conclusion n'est pas vraie, cela signifie que dans tout voisinage de x^*, on peut trouver un vecteur x différent de x^* et vérifiant $F(x) = 0$. Nous en déduisons qu'il existe une suite $\{x^k\}$ de $\mathbf{R}^n$, convergeant vers x^* telle que $x^k \neq x^*$ et $F(x^k) = 0$ pour tout k.

Définissons $\{h^k\} = \left\{ \dfrac{(x^k - x^*)}{\| x^k - x^* \|} \right\}$. Cette suite est bornée: nous pouvons donc supposer, en prenant si nécessaire une sous-suite, qu'elle converge vers un certain h de $\mathbf{R}^n$, où $\| h \| \neq 0$ (car $\| h \| = 1$).

Puisque F est semi-lisse, le théorème 0.2, point (ii), assure que

$$\lim_{\substack{h^k \to h \\ \|x^k - x^*\| \to 0}} \frac{F(x^* + \|x^k - x^*\| \, h^k) - F(x^*)}{\| x^k - x^* \|} = F'(x^*; h),$$

où nous avons identifié, dans le théorème, h' à h^k et t à $\| x^k - x^* \|$.

Nous obtenons alors que $F'(x^*; h) = 0$, car

$$F'(x^*; h) = \lim_{\substack{h^k \to h \\ \|x^k - x^*\| \to 0}} \frac{F(x^* + \|x^k - x^*\| \, h^k) - F(x^*)}{\| x^k - x^* \|}$$

$$= \lim_{k \to \infty} \frac{F(x^* + x^k - x^*) - F(x^*)}{\| x^k - x^* \|}$$

$$= 0, \qquad \text{puisque } F(x^*) = F(x^k) = 0.$$

Ceci contredit l'hypothèse de BD-régularité et prouve donc que x^* est, localement, solution unique de l'équation $F(x) = 0$. □

CHAPITRE 1

Une version non lisse de la méthode de Newton

La méthode de Newton, d'itération

$$x^{k+1} = x^k - [F'(x^k)]^{-1} F(x^k), \qquad (1.1)$$

est une méthode classique utilisée pour résoudre l'équation non linéaire $F(x) = 0$, où $F: \mathbf{R}^n \to \mathbf{R}^n$ est une fonction continûment différentiable, c'est-à-dire une *fonction lisse.* Beaucoup d'autres méthodes servant à résoudre $F(x) = 0$ sont liées à celle-ci.

Supposons maintenant que F ne soit plus une fonction lisse, mais seulement une fonction localement lipschitzienne. Dans ce cas, la formule (1.1) ne peut plus être utilisée, mais nous pouvons la remplacer par

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \qquad (1.2)$$

où V_k appartient à $\delta F(x^k)$.

Dans ce chapitre, nous allons étudier cette version non lisse de la méthode de Newton. Nous montrerons dans un premier temps que des résultats de convergence locale sont vérifiés lorsque F est semi-lisse.
Ensuite, nous comparerons ces résultats avec une autre version non lisse de la méthode de Newton, où la dérivée de Fréchet en (1.1) est remplacée par la B-dérivée. Nous remarquerons que cette dernière méthode peut être considérée comme un cas spécial de la version utilisant le jacobien généralisé, (1.2).
Enfin, nous verrons une application possible de la méthode étudiée: le lagrangien augmenté d'un problème de programmation convexe non linéaire de classe C^2 possédant

des gradients semi-lisses, la méthode du lagrangien augmenté peut faire intervenir l'itération (1.2) lors de la résolution de programmes non linéaires.

1.1 Méthode basée sur le jacobien généralisé

Soit $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$ une fonction localement lipschitzienne. Notre motivation est de trouver la solution de

$$F(x) = 0. \qquad (1.3)$$

Pour cela, nous allons considérer une version non lisse de la méthode de Newton, donnée par:

$$x^{k+1} = x^k - V_k^{-1} F(x^k),$$

où $V_k \in \delta F(x^k)$.

1.2 Convergence de la méthode

Pour analyser la convergence de cette méthode, nous avons besoin d'une propriété satisfaite par les éléments de l'ensemble $\delta F(x)$.

Proposition 1.1

> Si tous les V appartenant à $\delta F(x)$ sont non singuliers,
> alors il existe un voisinage U_x de x et une constante c tels que pour tout y dans U_x
> et pour tout V dans $\delta F(y)$, V est non singulier et $\| V^{-1} \| \leq c$.

PREUVE:
Supposons par l'absurde que la thèse ne soit pas vérifiée. Alors, pour tout voisinage U_x de x et pour toute constante c, il existe un vecteur y dans U_x et une matrice V dans $\delta F(y)$ tels que soit V est singulière, soit $\| V^{-1} \| > c$.

Notons par $B(x, \varepsilon)$ la boule ouverte de centre x et de rayon ε, c'est-à-dire

$$B(x, \varepsilon) = \{ y \in \mathbf{R}^n : \| y - x \| < \varepsilon \}.$$

Pour tout k dans **N**, choisissons $U_x{}^k = B(x, 1/k)$ et $c_k = k$. Nous pouvons alors affirmer que

$$\forall k \in \mathbf{N}, \exists\, y^k \in B(x, 1/k) \text{ et } V_k \in \delta F(y^k) \text{ tels que}$$

$$\text{soit } V_k \text{ est singulière, soit } \| V_k^{-1} \| > k,$$

ou encore

$$\exists \{y^k\}, y^k \to x \text{ et } \exists \{V_k\} \text{ telle que } \forall k, V_k \in \delta F(y^k) \text{ et}$$

$$V_k \text{ est soit singulière, soit vérifie } \| V_k^{-1} \| > k.$$

La suite $\{V_k\}$ est donc de la forme suivante: ou bien une infinité de ses éléments sont singuliers, ou bien elle contient un sous-suite $\{V_i\}$ telle que $\{V_i^{-1}\}$ est non bornée, ou encore elle est réunion de ces deux cas.

Nous pouvons donc conclure qu'il existe une suite $\{y^i\}$ convergeant vers x et une suite $\{V_i\}$, V_i appartenant à $\delta F(y^i)$ pour tout i, telle que soit toutes les matrices V_i sont singulières, soit la suite $\{\| V_i^{-1} \|\}$ tend vers $+\infty$.

Comme F est localement lipschitzienne, δF est borné au voisinage de x. Par conséquent, la suite $\{V_i\}$ est elle aussi bornée et en passant si nécessaire à une sous-suite, nous pouvons affirmer que $\{V_i\}$ converge vers une certaine limite V. Ceci implique alors que V appartienne à $\delta F(x)$ (car le jacobien généralisé est un ensemble fermé) et que V soit singulière.

Nous obtenons ainsi une contradiction, car nous avons trouvé une matrice V singulière dans $\delta F(x)$ et par hypothèse, tous les éléments de $\delta F(x)$ sont non singuliers. La thèse est donc vérifiée. □

Le théorème suivant établit la convergence locale de la méthode. Il est à remarquer que l'hypothèse de non singularité des éléments de $\delta F(x^*)$ représente la version non lisse de la condition imposée pour la méthode de Newton classique dans le cas lisse, à savoir la non singularité de la dérivée.

Théorème 1.1 [Convergence locale]

> Soit x^* une solution de l'équation $F(x) = 0$,
> où $F: \mathbf{R}^n \to \mathbf{R}^n$ est une fonction localement lipschitzienne et semi-lisse en x^*.
> Supposons que tous les V dans $\delta F(x^*)$ soient non singuliers.
> Alors la méthode d'itération (1.2) est bien définie et converge vers x^* dans un voisinage de x^*.
> Si en outre F est semi-lisse d'ordre p en x^*,
> alors la convergence est d'ordre $p+1$.

PREUVE:

Par la proposition précédente (proposition 1.1), nous obtenons:

$$\exists\, \delta_1 > 0, \exists\, c \geq 0 \text{ tels que } \forall\, x \in \mathbf{R}^n \text{ vérifiant } \| x - x^* \| < \delta_1, \forall\, V \in \delta F(x),$$

$$V \text{ est non singulier et } \| V^{-1} \| \leq c. \tag{1.4}$$

Montrons qu'il existe un δ strictement positif, $\delta \leq \delta_1$, tel que pour tout k dans $\mathbf{N}$, les itérés x^k de (1.2) satisfont $\| x^k - x^* \| < \delta$. (1.5)

Dans ce cas, nous aurons $\| x^k - x^* \| < \delta_1$ pour tout k et par conséquent, pour tout V_k dans $\delta F(x^k)$, V_k sera non singulier et $\| V_k^{-1} \| \leq c$. L'égalité (1.2) sera donc bien définie pour tout k dans un voisinage de x^*.

Choix de δ strictement positif

Nous savons par (0.21) que si F est semi-lisse en x^*, alors pour tout x tendant vers x^*, $F(x) - F(x^*) - F'(x^*; x - x^*) = o(\| x - x^* \|)$. Ceci peut s'exprimer comme suit:

$$\forall\, \varepsilon > 0, \exists\, \delta_2 > 0 \text{ tel que}$$

$$\| x - x^* \| < \delta_2 \Rightarrow \| F(x) - F(x^*) - F'(x^*; x - x^*) \| \leq \varepsilon \| x - x^* \|. \tag{1.6}$$

D'autre part, le caractère semi-lisse de F en x^* implique aussi que (voir théorème 0.2, (iv)) pour tout x tendant vers x^*, pour tout V dans $\delta F(x)$,
$V(x - x^*) - F'(x^*; x - x^*) = o(\| x - x^* \|)$, c'est-à-dire

$$\forall\, \varepsilon > 0, \exists\, \delta_3 > 0 \text{ tel que} \tag{1.7}$$

$$\| x - x^* \| < \delta_3, V \in \delta F(x) \Rightarrow \| V(x - x^*) - F'(x^*; x - x^*) \| \leq \varepsilon \| x - x^* \|.$$

Soit $\delta = \min \{\delta_1, \delta_2, \delta_3\}$.

En choisissant $\varepsilon = \frac{1}{3c}$ dans (1.6) et (1.7), nous pouvons maintenant prouver (1.5) par récurrence:

• pour $k = 0$, choisir x^o tel que $\| x^o - x^* \| < \delta$. L'égalité (1.2) est alors bien définie dans ce cas, puisque $\delta \leq \delta_1$.

• supposons que $\| x^k - x^* \| < \delta$; montrons que $\| x^{k+1} - x^* \| < \delta$:

$$\begin{aligned}
\| x^{k+1} - x^* \| &= \| x^k - x^* - V_k^{-1} F(x^k) \| \\
&= \| - V_k^{-1} \{F(x^k) - F(x^*) - F'(x^*; x^k - x^*)\} \\
&\quad + V_k^{-1} \{V_k (x^k - x^*) - F'(x^*; x^k - x^*)\} \| \\
&\qquad\qquad \text{où } F(x^*) = 0 \text{ par hypothèse} \\
&\leq \| V_k^{-1} \{F(x^k) - F(x^*) - F'(x^*; x^k - x^*)\} \| \\
&\quad + \| V_k^{-1} \{V_k (x^k - x^*) - F'(x^*; x^k - x^*)\} \| \\
&\leq \| V_k^{-1} \| \big[\| F(x^k) - F(x^*) - F'(x^*; x^k - x^*) \| \\
&\quad + \| V_k (x^k - x^*) - F'(x^*; x^k - x^*) \|\big]. \qquad (1.8)
\end{aligned}$$

L'hypothèse de récurrence ($\| x^k - x^* \| < \delta$) nous permet alors d'utiliser (1.4), (1.6) et (1.7) afin de majorer l'expression (1.8). Nous obtenons ainsi:

$$\begin{aligned}
\| x^{k+1} - x^* \| &\leq (1.8) \\
&\leq c \left[\frac{1}{3c} \| x^k - x^* \| + \frac{1}{3c} \| x^k - x^* \|\right] = \frac{2}{3} \| x^k - x^* \| \\
&< \frac{2}{3} \delta < \delta.
\end{aligned}$$

La preuve de (1.5) est donc achevée.

Nous avons en outre obtenu la convergence locale de la suite $\{x^k\}$ vers x^*. En effet, nous avons vu ci-dessus que

$$\| x^{k+1} - x^* \| \leq \frac{2}{3} \| x^k - x^* \| \quad \text{pour tout } k.$$

Par conséquent, il existe un scalaire γ, avec $0 < \gamma < 1$, tel que

$$\| x^{k+1} - x^* \| < \gamma \| x^k - x^* \| \quad \text{pour tout } k.$$

La preuve est analogue lorsque F est semi-lisse d'ordre p en x^*. □

Nous avons établi ci-dessus la convergence locale de la méthode étudiée. Il est à remarquer que sous les hypothèses citées, le taux de convergence de la suite des itérés $\{x^k\}$ est superlinéaire.
En effet, nous savons que si F est semi-lisse en x^*, alors pour toute suite $\{x^k\}$ convergeant vers x^* et pour toute matrice V_k dans $\delta F(x^k)$,

$$F(x^k) - F(x^*) - F'(x^*; x^k - x^*) = o(\| x^k - x^* \|)$$

et

$$V_k (x^k - x^*) - F'(x^*; x^k - x^*) = o(\| x^k - x^* \|).$$

Par conséquent,

$$\begin{aligned} \| x^{k+1} - x^* \| &\leq \| V_k^{-1} \| \big[\| F(x^k) - F(x^*) - F'(x^*; x^k - x^*) \| \\ &\qquad + \| V_k (x^k - x^*) - F'(x^*; x^k - x^*) \| \big] \qquad \text{(voir (1.8))} \\ &\leq c \big[o(\| x^k - x^* \|) + o(\| x^k - x^* \|) \big] \\ &= o(\| x^k - x^* \|), \end{aligned}$$

la suite converge donc bien de manière superlinéaire.

Etudions à présent un autre théorème de convergence, qui permettra d'estimer l'erreur à chaque itération.

Théorème 1.2

Soit F: $\mathbf{R}^n \rightarrow \mathbf{R}^n$ une fonction localement lipschitzienne et semi-lisse sur $S = \{ x \in \mathbf{R}^n : \| x - x^o \| \leq r \}$.
Supposons que pour tout V dans $\delta F(x)$, pour tout x, y dans S,

$$V \text{ est non singulier, } \| V^{-1} \| \leq \beta,$$

$$\| V (y - x) - F'(x; y - x) \| \leq \gamma \| y - x \| \text{ et}$$

$$\| F(y) - F(x) - F'(x; y - x) \| \leq \delta \| y - x \|,$$

où $\alpha = \beta (\gamma + \delta) < 1$ et $\beta \| F(x^o) \| \leq r (1 - \alpha)$.

Alors, les itérés de (1.2) restent tous dans S et convergent vers la solution unique x^* de l'équation $F(x) = 0$, dans S.
En outre, l'estimateur de l'erreur,

$$\| x^k - x^* \| \leq [\alpha / (\alpha - 1)] \| x^k - x^{k-1} \|, \tag{1.9}$$

est valable pour tout k dans **N**.

PREUVE:
Nous allons d'abord vérifier, par récurrence, que les itérés de (1.2) restent bien dans S:

- $x^1 \in S$, car:

$$\| x^1 - x^o \| = \| V_o^{-1} F(x^o) \| \qquad \text{par (1.2)}$$

$$\leq \beta \| F(x^o) \| \leq r (1 - \alpha), \qquad \text{par hypothèse.}$$

Comme α est strictement inférieur à 1, nous obtenons bien $\| x^1 - x^o \| \leq r$.

- Si $x^o, x^1, .., x^k$ appartiennent à S, alors nous avons successivement:

$$\| x^{k+1} - x^k \| = \| V_k^{-1} F(x^k) \| \qquad \text{par (1.2)}$$

$$\leq \beta \| F(x^k) \| \qquad \text{car } V_k \in \delta F(x^k) \text{ et } x^k \in S$$

$$\leq \beta \| F(x^k) - F(x^{k-1}) - F'(x^{k-1}; x^k - x^{k-1}) \|$$

$$+ \beta \| V_{k-1} (x^k - x^{k-1}) - F'(x^{k-1}; x^k - x^{k-1}) \|$$

$$\leq \beta \delta \| x^k - x^{k-1} \| + \beta \gamma \| x^k - x^{k-1} \| \qquad \text{par hypothèse}$$

$$= \beta (\delta + \gamma) \| x^k - x^{k-1} \|$$

$$= \alpha \| x^k - x^{k-1} \| \leq \ldots \leq \alpha^k \| x^1 - x^o \|$$

$$\leq \alpha^k r (1 - \alpha).$$

Nous pouvons en déduire que x^{k+1} appartient à S, car

$$\| x^{k+1} - x^o \| \leq \sum_{j=0}^{k} \| x^{j+1} - x^j \|$$

$$\leq \sum_{j=0}^{k} \alpha^j r (1 - \alpha) \leq \sum_{j=0}^{\infty} \alpha^j r (1 - \alpha)$$

$$= r, \qquad \text{puisque } 0 \leq \alpha < 1.$$

Etudions maintenant la convergence de la suite des itérés $\{x^k\}$. Pour k et p arbitraires dans **N**, nous obtenons les inégalités suivantes:

$$\| x^{k+p+1} - x^k \| \leq \sum_{j=k}^{k+p} \| x^{j+1} - x^j \| \leq \sum_{j=k}^{k+p} \alpha^j r (1 - \alpha)$$

$$= r (1 - \alpha) \frac{\alpha^k - \alpha^{k+p+1}}{1 - \alpha} \leq r \alpha^k. \qquad (1.10)$$

Comme α est strictement inférieur à 1, il découle de (1.10) que la suite $\{x^k\}$ est de Cauchy, donc convergente vers un certain x^*. Puisque toute la suite est dans S, S étant fermé par définition, la limite x^* appartient aussi à S.

Ensuite, montrons que x^* est solution de $F(x) = 0$ et que cette solution est unique dans S. F étant une fonction localement lipschitzienne dans S, la suite $\{\| V_k \|\}$ est uniformément bornée. Nous pouvons en déduire que x^* vérifie $F(x^*) = 0$:

$$0 \leq \| F(x^*) \| = \lim_{k \to \infty} \| F(x^k) \| \leq \lim_{k \to \infty} \| V_k \| \| x^{k+1} - x^k \| = 0,$$

car $\| V_k \|$ est bornée et $\| x^{k+1} - x^k \|$ converge vers zéro.

Supposons à présent qu'il existe un y^* dans S, tel que $F(y^*) = 0$. Prenons V^* dans $\delta F(x^*)$. Nous avons alors:

$$\| y^* - x^* \| = \| V^{*-1} V^* (y^* - x^*) \| \leq \beta \| V^* (y^* - x^*) \|$$

$$\leq \beta \| V^* (y^* - x^*) - F'(x^*; y^* - x^*) \|$$

$$+ \beta \| F(y^*) - F(x^*) - F'(x^*; y^* - x^*) \|,$$

où $F(x^*) = F(y^*) = 0$,

$$\leq \beta (\delta + \gamma) \| y^* - x^* \| = \alpha \| y^* - x^* \|,$$

c'est-à-dire

$$\| y^* - x^* \| \leq \alpha \| y^* - x^* \|, \qquad \text{avec } 0 \leq \alpha < 1.$$

Par conséquent, $\| y^* - x^* \| \leq 0$ et nous obtenons ainsi $y^* = x^*$.
La solution x^* de $F(x) = 0$ est donc bien unique dans S.

Finalement, estimons l'erreur pour tout k dans **N**:

$$\| x^{k+p+1} - x^k \| \leq \sum_{j=k}^{k+p} \| x^{j+1} - x^j \| \leq \sum_{j=0}^{p} \alpha^{j+1} \| x^k - x^{k-1} \|$$

$$= \frac{\alpha - \alpha^{p+2}}{1 - \alpha} \| x^k - x^{k-1} \|$$

$$\leq [\alpha / (\alpha - 1)] \| x^k - x^{k-1} \|.$$

Il suffit enfin de faire tendre p vers $+\infty$ pour obtenir l'expression (1.9). □

Remarque

Pour obtenir α strictement inférieur à 1, il est nécessaire que γ et δ soient petits; cette condition peut être considérée comme une forme globale de (0.7) et (0.21).
La condition de petitesse de gamma peut être vérifiée lorsque le diamètre de $\delta F(x)$ est petit pour tout x dans S.

1.3 Comparaison avec la méthode de Newton basée sur la B-dérivée

La méthode de Newton utilisant la B-dérivée est définie par:

$$x^{k+1} = x^k + d^k \tag{1.11}$$

et

$$F(x^k) + BF(x^k)\, d^k = 0. \tag{1.12}$$

En général, l'égalité (1.12) est un système d'équations non linéaires. Le résoudre n'est pas toujours une tâche triviale. J. S. Pang (voir [5]) a suggéré de le faire de manière inexacte, mais cette résolution n'est pas notre but ici. Nous allons prouver que la méthode définie ci-dessus peut être considérée comme un cas spécial de la méthode proposée dans le paragraphe 1.1 de ce chapitre (voir proposition 1.2). Nous montrerons aussi que le théorème de convergence locale établi pour la méthode basée sur le jacobien généralisé est plus puissant que les résultats obtenus pour la méthode définie par le système (1.11)–(1.12).

Proposition 1.2

Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne en x.
Supposons que tout V dans $\delta F(x)$ soit non singulier.
Alors, dans un certain voisinage de x et pour un choix convenable de V_k dans $\delta F(x^k)$, les itérés générés par (1.11) et (1.12) coïncident avec ceux obtenus par (1.2).

PREUVE:
Nous avons vu dans le chapitre préliminaire que la B-dérivée d'une fonction en x était identique à la dérivée directionnelle de cette fonction en x, c'est-à-dire que

$$BF(x)\, h = F'(x; h)$$

pour toute direction h, lorsqu'une de ces dérivées existe.
Dès lors, nous pouvons remplacer l'équation (1.12) par

$$F(x^k) + F'(x^k; d^k) = 0.$$

En outre, la proposition 0.2 (chapitre préliminaire) établit que pour toute direction d^k, il existe une matrice V_k dans $\delta F(x^k)$ telle que

$$F'(x^k; d^k) = V_k d^k.$$

D'après la proposition 1.1, V_k est inversible lorsque x^k est dans un voisinage adéquat de x. Nous avons ainsi obtenu qu'il existe un voisinage de x tel que si x^k est dans ce voisinage et V_k dans $\delta F(x^k)$, alors V_k^{-1} existe et (1.12) devient:

$$F(x^k) + V_k d^k = 0,$$

d'où

$$d^k = - V_k^{-1} F(x^k).$$

Par conséquent, les itérés générés par (1.11) et (1.12) sont les mêmes que ceux obtenus par (1.2), pour un choix convenable de V_k dans $\delta F(x^k)$. □

Proposition 1.3

Si la B-dérivée de F, BF(•), est une fonction lipschitzienne en x, alors F est semi-lisse d'ordre 1 en x.

PREUVE:

Par définition, BF(•) est lipschitzienne en x équivaut à dire que:

$$\exists\, U_x \text{ voisinage de } x,\ \exists\, L \geq 0 \text{ tels que } \forall\, y \in U_x,\ \forall\, h \in \mathbf{R}^n \text{ avec } \| h \| = 1,$$

$$\| (BF(y) - BF(x))\, h \| \leq L \| y - x \|. \qquad (0.23)$$

Prenons $y = x + h$ dans U_x et V dans $\delta F(x+h)$.

La définition du jacobien généralisé (voir (0.1)) nous donne:

$$\delta F(x+h)\, h = \text{co} \{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} F'(x_i + h; h) \}. \qquad (1.13)$$

En effet, $\delta F(x+h)\, h = \text{co} \{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} JF(x_i + h) \}\, h$

$$= \mathrm{co}\ \{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} JF(x_i + h)\, h \}$$

$$= \mathrm{co}\ \{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} F'(x_i + h; h) \}.$$

Comme Vh appartient à $\delta F(x+h)h$, en vertu du théorème de Carathéodory nous déduisons de (1.13) que Vh peut s'exprimer comme

$$Vh = \sum_{k=0}^{m} \lambda_k X^{(k)}$$

où, pour tout k, $X^{(k)} \in \{ \lim_{\substack{x_i \to x \\ x_i + h \in D_F}} F'(x_i + h; h) \}$ et $\lambda_k \geq 0$, tels que $\sum_{k=0}^{m} \lambda_k = 1$,

c'est-à-dire

$$Vh = \sum_{k=0}^{m} \lambda_k X^{(k)}$$

où, pour tout k, $X^{(k)} = \lim_{\substack{x_i^{(k)} \to x \\ x_i^{(k)} + h \in D_F}} F'(x_i^{(k)} + h; h)$ et $\lambda_k \geq 0$, tels que $\sum_{k=0}^{m} \lambda_k = 1$,

ou encore

$$Vh = \sum_{k=0}^{m} \lambda_k \lim_{\substack{x_i^{(k)} \to x \\ x_i^{(k)} + h \in D_F}} F'(x_i^{(k)} + h; h). \tag{1.14}$$

Nous obtenons alors qu'il existe une suite $\{x_i^{(p)}\}$, $i \in \mathbf{N}$, convergeant vers x, telle que $x_i^{(p)} + h$ appartienne à D_F pour tout i et

$$\lim_{i \to \infty} \| F'(x_i^{(p)} + h; h) - F'(x; h) \| \geq \| Vh - F'(x; h) \|, \tag{1.15}$$

car nous avons successivement

$$\| Vh - F'(x; h) \| = \| \sum_{k=0}^{m} \lambda_k [\lim_{\substack{x_i^{(k)} \to x \\ x_i^{(k)} + h \in D_F}} F'(x_i^{(k)} + h; h) - F'(x; h)] \|$$

$$\leq \sum_{k=0}^{m} \lambda_k \, \| \lim_{\substack{x_i^{(k)} \to x \\ x_i^{(k)}+h \in D_F}} F'(x_i^{(k)} + h; h) - F'(x; h) \|$$

$$\leq \sum_{k=0}^{m} \lambda_k \sup_k \{ \| \lim_{\substack{x_i^{(k)} \to x \\ x_i^{(k)}+h \in D_F}} F'(x_i^{(k)} + h; h) - F'(x; h) \| \}$$

$$\leq 1 \lim_{\substack{x_i^{(p)} \to x \\ x_i^{(p)}+h \in D_F}} \| F'(x_i^{(p)} + h; h) - F'(x; h) \|,$$

où l'indice p correspond à l'indice du supremum et la suite $\{x_i^{(p)}\}$ tend vers x lorsque i tend vers l'infini, ce qui correspond bien à (1.15).

Pour conclure, nous déduisons de (0.23) et (1.15) que

$$\| Vh - F'(x; h) \| \leq \lim_{i \to \infty} \| F'(x_i^{(p)} + h; h) - F'(x; h) \| \qquad \text{par (1.15)}$$

$$= \| (BF(x + h) - BF(x)) \, h \| \frac{\| h \|}{\| h \|}$$

$$\leq L \, \| x + h - x \| \; \| h \| \qquad \text{par (0.23)}$$

$$= L \, \| h \|^2.$$

Ceci prouve que F est semi-lisse d'ordre 1, puisque nous avons obtenu que

$$\| Vh - F'(x; h) \| = O(\| h \|^2). \qquad \square$$

Conclusion

J. S. Pang, dans son article (voir [5]), avait établi un théorème de convergence locale, pour la méthode de Newton basée sur la B-dérivée. Nous allons montrer que le théorème 1.1 démontré plus haut apporte un résultat plus puissant.

Théorème 1.3 **[Pang, 1989]**

> Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction lipschitzienne et B-différentiable dans un voisinage d'une solution x^* de l'équation $F(x) = 0$.
> Supposons que la dérivée de Fréchet $F'(x^*)$ existe, est forte et non singulière.
> Alors, il existe un voisinage U_{x^*} de x^* tel que pour tout vecteur initial x^o dans U_{x^*}, les itérés x^k soient bien définis par (1.1.11) et (1.1.12), restent dans U_{x^*} et convergent vers x^*.
> En outre, si BF(•) est lipschitzienne en x^*, le taux de convergence est quadratique.

□

En vue de la proposition 1.2, nous constatons que les théorèmes 1.1 et 1.2 sont encore valables pour les itérés générés par la méthode utilisant la B-dérivée, tant que l'équation (1.12) peut être résolue.
Par le corollaire 0.4, l'existence de la dérivée forte de Fréchet de F en x^* implique que F soit semi-lisse en x^*. De plus, si BF(•) est lipschitzienne en x^*, alors F est également semi-lisse d'ordre 1 en ce point.
Le théorème 1.1 est donc plus puissant que celui de Pang, puisqu'une fonction semi-lisse peut ne pas avoir de dérivée de Fréchet en un point.

1.4 Application: le lagrangien augmenté

Considérons le programme non linéaire suivant, de classe C^2, noté (PNL):

$$\left|\begin{array}{lll} \text{minimiser } f_o(x) & & \\ \text{sous contraintes} & & \\ \quad f_i(x) = 0 & \text{pour } i = 1, 2,.., p & \\ \quad f_i(x) \leq 0 & \text{pour } i = p+1,.., m, & \end{array}\right. \qquad \text{(PNL)}$$

où $x \in \mathbf{R}^n$ et $f_i \in C^2$ pour $i = 0, 1,.., m$.

Le lagrangien augmenté de ce programme (PNL) est donné par:

$$L_A(x, y) = f_0(x) + \sum_{i=1}^{p} h(r, f_i(x), y_i) + \sum_{i=p+1}^{m} \phi(r, f_i(x), y_i),$$

où $r > 0$,
$h(r, f_i(x), y_i) = y_i f_i(x) + \frac{1}{2} r f_i(x)^2$ et

$$\phi(r, f_i(x), y_i) = \begin{cases} y_i f_i(x) + \frac{1}{2} r f_i(x)^2 & \text{si } y_i + r f_i(x) \geq 0, \\ -\frac{1}{2r} y_i^2 & \text{si } y_i + r f_i(x) \leq 0. \end{cases}$$

Théorème 1.4

Si les hypothèses ci-dessus sont vérifiées,
alors $L_A(\bullet, \bullet)$ est de classe C^1,
$\nabla L_A(\bullet, \bullet)$ est semi-lisse sur les surfaces d'équation $y_i + r f_i(x) = 0$, $i = p+1,.., m$ et lisse ailleurs.

PREUVE:
Fixons i appartenant à $\{p+1,.., m\}$ et r strictement positif, arbitraires.
Notons $\phi(r, f_i(x), s)$ par $\eta(x, s)$.
Nous obtenons ainsi

$$\eta(x, s) = \begin{cases} s f_i(x) + \frac{1}{2} r f_i(x)^2 & \text{si } s + r f_i(x) \geq 0, \\ -\frac{1}{2r} s^2 & \text{si } s + r f_i(x) \leq 0. \end{cases}$$

Pour prouver le théorème, il suffit alors montrer que η est de classe C^1 et que $\nabla\eta$ est semi-lisse sur la surface d'équation $s + r f_i(x) = 0$ et lisse ailleurs.

En effet, la fonction h ne posera pas de problème puisqu'elle est, par construction, de classe C^2.

Calculons $\nabla\eta(x, s)$:

$$\nabla\eta(x, s) = \begin{cases} \begin{pmatrix} (s + r f_i(x))\nabla f_i(x) \\ f_i(x) \end{pmatrix} & \text{si } s + r f_i(x) \geq 0, \\ \begin{pmatrix} 0 \\ -s/r \end{pmatrix} & \text{si } s + r f_i(x) \leq 0. \end{cases}$$

Il est alors clair que η est de classe C^1 et que $\nabla\eta$ est localement lipschitzienne, en vue des hypothèses.

Il reste à montrer que $\nabla\eta$ est semi-lisse sur la surface d'équation $s + r\, f_i(x) = 0$ et lisse ailleurs.

D'une part, $\nabla^2\eta(x, s)$ est lisse, excepté sur la surface citée ci-dessus, car

$$\nabla^2\eta(x, s) = \begin{cases} \begin{pmatrix} r\,(\nabla f_i(x))(\nabla f_i(x))^T + (s + r\, f_i(x))\nabla^2 f_i(x) & \nabla f_i(x) \\ (\nabla f_i(x))^T & 0 \end{pmatrix} & \text{si } s + r\, f_i(x) > 0 \\ \begin{pmatrix} 0 & 0 \\ 0 & -\frac{1}{r} \end{pmatrix} & \text{si } s + r\, f_i(x) < 0. \end{cases} \qquad (1.16)$$

Ceci montre que $\nabla\eta(\bullet, \bullet)$ est lisse, excepté sur la surface d'équation $s + r\, f_i(x) = 0$.

D'autre part, prenons un point $(\bar{x}, \bar{s})$ sur cette surface, c'est-à-dire un point $(\bar{x}, \bar{s})$ vérifiant

$$\bar{s} + r\, f_i(\bar{x}) = 0. \qquad (1.17)$$

Nous allons prouver que $\nabla\eta(\bullet, \bullet)$ est semi-lisse en $(\bar{x}, \bar{s})$.
Soit (h, α) un vecteur de $\mathbf{R}^n \times \mathbf{R}$. Il suffit alors de montrer que

$$\lim_{\substack{h' \to h, \alpha' \to \alpha \\ t \downarrow 0}} \left\{ V \begin{pmatrix} h' \\ \alpha' \end{pmatrix} : V \in \delta\nabla\eta(\bar{x} + t\, h', \bar{s} + t\, \alpha') \right\} \qquad (1.18)$$

existe, puisque nous avons vu plus haut que $\nabla\eta$ était localement lipschitzienne.

La condition (1.18) peut encore s'exprimer de la manière suivante:

si $\{h^k\}$, $\{\alpha^k\}$ et $\{t_k\}$, $k \in \mathbf{N}$, sont des suites telles que
$h^k \to h$, $\alpha^k \to \alpha$ et $t_k \downarrow 0$ quand $k \to \infty$, la limite

$$\lim_{k\to\infty} \left\{ V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix} : V \in \delta\nabla\eta(\overline{x} + t\,h^k, \overline{s} + t_k\alpha^k) \right\} \text{ existe.} \tag{1.19}$$

Supposons d'abord que l'expression

$$\overline{s} + t_k\,\alpha^k + r\,f_i(\overline{x} + t_k\,h^k) \tag{1.20}$$

garde le même signe quand k tend vers l'infini. Dans ce cas, en utilisant (1.16) il est facile de voir que la limite (1.19) existe.

Supposons ensuite que l'expression (1.20) change de signe lorsque k converge vers l'infini.
Alors, il existe des sous-suites $\{h^{k_j}\}$ de $\{h^k\}$ et $\{\alpha^{k_j}\}$ de $\{\alpha^k\}$ convergeant respectivement vers h et α , ainsi qu'une sous-suite $\{t_{k_j}\}$ de $\{t_k\}$ décroissant vers zéro quand j tend vers l'infini, telles que

$$\overline{s} + t_{k_j}\,\alpha^{k_j} + r\,f_i(\overline{x} + t_{k_j}\,h^{k_j}) = 0. \tag{1.21}$$

En soustrayant (1.17) à (1.21), nous obtenons:

$$t_{k_j}\,\alpha^{k_j} + r\,\{f_i(\overline{x} + t_{k_j}\,h^{k_j}) - f_i(\overline{x})\} = 0,$$

c'est-à-dire

$$\alpha^{k_j} + \frac{r\,\{f_i(\overline{x} + t_{k_j}\,h^{k_j}) - f_i(\overline{x})\}}{t_{k_j}} = 0.$$

Comme f_i est de classe C^2, le passage à la limite lorsque j tend vers l'infini dans l'expression ci-dessus nous donne:

$$\alpha + r\,f_i'(\overline{x}; h) = 0,$$

ou encore

$$\alpha + r\,\nabla f_i(\overline{x})^T h = 0. \tag{1.22}$$

Maintenant, considérons les suites $\{h^k\}$, $\{\alpha^k\}$ et $\{t_k\}$.
Soit V dans $\delta\nabla\eta(\overline{x} + t_k\,h^k, \overline{s} + t_k\,\alpha^k)$.
Si l'expression $\overline{s} + t_k\,\alpha^k + r\,f_i(\overline{x} + t_k\,h^k)$ est négative, alors, par (1.16),

$$V = \begin{pmatrix} 0 & 0 \\ 0 & \frac{-1}{r} \end{pmatrix}.$$

En prenant si nécessaire une sous-suite, on peut affirmer que

$$V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix} = \begin{pmatrix} 0 \\ -\alpha^k/r \end{pmatrix} \to \begin{pmatrix} 0 \\ -\alpha/r \end{pmatrix} \tag{1.23}$$

quand k converge vers l'infini.

Si l'expression $\bar{s} + t_k \alpha^k + r f_i(\bar{x} + t_k h^k)$ est positive, alors, par (1.16),

$$V = \begin{pmatrix} r(\nabla f_i(\bar{x} + t_k h^k))(\nabla f_i(\bar{x} + t_k h^k))^T + (\bar{s} + t_k\alpha^k + r f_i(\bar{x} + t_k h^k))\nabla^2 f_i(\bar{x} + t_k h^k) & \nabla f_i(\bar{x} + t_k h^k) \\ (\nabla f_i(\bar{x} + t_k h^k))^T & 0 \end{pmatrix}$$

Donc, en prenant si nécessaire une sous-suite, nous obtenons

$$V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix} \to \begin{pmatrix} (\nabla f_i(\bar{x})(r(\nabla f_i(\bar{x})^T h + \alpha) + (\bar{s} + r f_i(\bar{x}))\nabla^2 f_i(\bar{x})h \\ (\nabla f_i(\bar{x})^T h \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ -\alpha/r \end{pmatrix}, \tag{1.24}$$

où la dernière égalité est due aux relations (1.17) et (1.22).

Par conséquent, (1.23) et (1.24) nous permettent d'affirmer que l'expression $V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix}$ converge vers la même limite dans les deux situations étudiées ci-dessus.

Lorsque $\bar{s} + t_k \alpha^k + r f_i(\bar{x} + t_k h^k)$ vaut zéro, V est une combinaison convexe des deux cas précédents (par (1.16) et la définition du jacobien généralisé, (0.1)). D'où $V\begin{pmatrix} h^k \\ \alpha^k \end{pmatrix}$ converge aussi vers la même limite.

En conclusion, la limite en (1.19) existe, ce qui montre que $\nabla\eta(\bullet, \bullet)$ est semi-lisse en $(\overline{x}, \overline{s})$. La preuve est ainsi achevée. $\square$

L'étape de base de la méthode du lagrangien augmenté pour résoudre le programme (PNL) est de minimiser $L_A(x, y)$. Par conséquent, la version de la méthode de Newton proposée en (1.2) et le théorème 1.1 peuvent être appliqués ici.

CHAPITRE 2

Amélioration de la méthode

Dans le chapitre précédent, nous avons étudié la méthode d'itération

$$x^{k+1} = x^k - V_k^{-1} F(x^k),$$

où V_k appartenait à $\delta F(x^k)$, afin de résoudre l'équation $F(x) = 0$ pour une fonction F localement lipschitzienne. La convergence locale superlinéaire a été prouvée, sous certaines conditions (voir théorème 1.1).

Nous proposerons ici une modification de cette méthode, en prenant V_k non pas dans $\delta F(x^k)$, mais dans l'ensemble $\delta_B F(x^k) = \{\lim_{\substack{x_i \to x^k \\ x_i \in D_F}} JF(x_i)\}$, afin d'augmenter "sa résolvabilité".

Nous étudierons ensuite la convergence locale de la nouvelle méthode, en affaiblissant les hypothèses du théorème 1.1: au lieu de supposer que pour un zéro x^* de F, toutes les matrices V de $\delta F(x^*)$ soient non singulières, nous imposerons seulement que cette condition soit satisfaite par les matrices de $\delta_B F(x^*)$. Nous montrerons également que sous cette hypothèse, la suite $\{F(x^k)\}$ vérifie

$$\lim_{k\to\infty} \frac{\| F(x^{k+1}) \|}{\| F(x^k) \|} = 0,$$

à condition que le vecteur initial x^o soit proche de x^* et que F soit semi-lisse en x^*.

2.1 La méthode améliorée

Notons $\delta_B F(x)$ l'ensemble $\{\lim_{\substack{x_i \to x \\ x_i \in D_F}} JF(x_i)\}$.

D'après la définition du jacobien généralisé, nous voyons que $\delta F(x) = \text{co}\ \delta_B F(x)$.

Dans le premier chapitre, nous avons étudié une version de la méthode de Newton basée sur le jacobien généralisé (voir (1.2)), afin de résoudre l'équation $F(x) = 0$. Le théorème 1.1 assurait la convergence locale de cette méthode. Dans ce théorème, nous supposions que toutes les matrices V appartenant à $\delta F(x^*)$ étaient non singulières, où x^* était un zéro de F. En réalité, cette hypothèse est trop forte. En effet, considérons la fonction f définie par $f(x) = | x |$: la méthode converge trivialement dans cet exemple, alors que le jacobien généralisé en zéro, $\delta F(0)$, contient zéro.

Nous suggérons ici une modification de la méthode, afin de relâcher la condition citée. L'itération principale proposée est la suivante:

$$x^{k+1} = x^k - V_k^{-1} F(x^k),$$

avec V_k appartenant à $\delta_B F(x^k)$. (2.1)

La différence avec la méthode proposée en (1.2) est donc que cette fois, nous prenons les matrices V_k dans $\delta_B F(x^k)$ et non dans l'enveloppe convexe de cet ensemble.

2.2 Convergence de la méthode améliorée

Définition 2.1

Une fonction $F: \mathbf{R}^n \to \mathbf{R}^m$ est fortement BD-régulière en x si toutes les matrices V de $\delta_B F(x)$ sont non singulières. (2.2)

Nous allons d'abord prouver une affirmation analogue à la proposition 1.1, établissant une propriété vérifiée par les matrices de $\delta_B F(x)$.

Proposition 2.1

> Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne.
> Si F est fortement BD-régulière en x, alors il existe un voisinage U de x et une constante c tels que pour tout y dans U, pour tout V dans $\delta_B F(y)$,
> V est non singulière et $\| V^{-1} \| \leq c$. (2.3)

PREUVE:

Dans un premier temps, montrons qu'il existe un voisinage U de x et une constante c tels que pour tout y dans $D_F \cap U$, JF(y) est non singulier et $\| [JF(y)]^{-1} \| \leq c$. (2.4)

Supposons par l'absurde que cela ne soit pas le cas. Alors, par un raisonnement analogue à celui utilisé dans la preuve de la proposition 1.1 (chapitre 1), nous obtenons qu'il existe une suite $\{y^k\}$ dans $D_F \cap U$, convergeant vers y, telle que soit $JF(y^k)$ est une matrice singulière pour tout k, soit $\| [JF(y^k)]^{-1} \|$ tend vers l'infini.

Comme F est localement lipschitzienne, δF est borné au voisinage de x. Par conséquent, la suite $\{JF(y^k)\}$ est elle aussi bornée et en passant si nécessaire à une sous-suite, nous pouvons affirmer que $\{JF(y^k)\}$ converge vers une certaine limite V. Ceci entraîne alors que V appartienne à $\delta_B F(x)$ et que V soit une matrice singulière.

Nous avons donc trouvé une matrice V singulière dans $\delta_B F(x)$: cette dernière affirmation contredit l'hypothèse de BD-régularité et l'assertion (2.4) est ainsi prouvée.

D'après (2.4) et la définition de $\delta_B F(x)$, nous pouvons conclure que pour tout y dans U et pour tout V dans $\delta_B F(y)$, V est non singulière et $\| V^{-1} \| \leq c$. (2.5)

En effet, soit y un vecteur arbitraire de U: y appartient ou non à D_F.
L'affirmation (2.5) a été prouvée ci-dessus pour y dans D_F (voir (2.4)). Supposons alors que y n'appartient pas à D_F. Par définition de $\delta_B F(y)$, il existe une suite $\{y^i\}$ dans D_F telle que V puisse s'exprimer de la manière suivante:

$$V = \lim_{\substack{y^i \to y \\ y^i \in D_F}} JF(y^i).$$

Comme à partir d'un certain indice, tous les éléments de la suite $\{y^i\}$ se trouvent dans U, nous obtenons par (2.4) que, pour tout indice i assez grand, $JF(y^i)$ est non singulière et $\| [JF(y^i)]^{-1} \| \leq c$. En faisant tendre i vers l'infini, il est facile de voir que le résultat (2.5) est vérifié.

La preuve de la proposition est ainsi terminée. □

Etudions maintenant la convergence locale de la méthode proposée: le théorème suivant assure qu'au voisinage d'une solution de $F(x) = 0$, la méthode converge de manière superlinéaire et est bien une méthode de descente; sous des hypothèses plus fortes, le taux de convergence est quadratique.

Théorème 2.1 [Convergence locale]

Soient $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et x^* une solution de l'équation $F(x) = 0$.
Supposons que F soit semi-lisse et fortement BD-régulière en x^*.
Alors, au voisinage de x^*, la méthode d'itération (2.1) est bien définie et converge vers x^* de manière superlinéaire. (2.7)

De plus, si $F(x^k)$ est différent de zéro pour tout k,
alors

$$\lim_{k\to\infty} \frac{\| F(x^{k+1}) \|}{\| F(x^k) \|} = 0, \tag{2.8}$$

pourvu que le vecteur initial x^0 soit assez proche de x^*.

En outre, si F est différentiable dans toutes les directions au voisinage de x^* et semi-continue de degré 2 en x^*, alors le taux de convergence de (2.1) est quadratique.

PREUVE:
Remarque: pour la première partie de la preuve, nous allons appliquer le même raisonnement que pour la preuve du théorème 1.1.
Par la proposition précédente (proposition 2.1), nous obtenons:

$$\exists\, \delta_1 > 0, \exists\, c \geq 0 \text{ tels que } \forall\, x \in \mathbf{R}^n \text{ vérifiant } \| x - x^* \| < \delta_1, \forall\, V \in \delta_B F(x),$$

$$V \text{ est non singulier et } \| V^{-1} \| \leq c. \tag{2.9}$$

Montrons qu'il existe un δ strictement positif, $\delta \le \delta_1$, tel que pour tout k dans **N**, les itérés x^k de (2.1) satisfont $\| x^k - x^* \| < \delta$. (2.10)

Dans ce cas, nous aurons $\| x^k - x^* \| < \delta_1$ pour tout k et par conséquent, pour tout V_k dans $\delta_B F(x^k)$, V_k sera non singulier et $\| V_k^{-1} \| \le c$. L'égalité en (2.1) sera donc bien définie pour tout k dans un voisinage de x^*.

<u>Choix de δ strictement positif</u>

Nous savons par (0.21) que si F est semi-lisse en x^*, alors pour tout x tendant vers x^*, $F(x) - F(x^*) - F'(x^*; x - x^*) = o(\| x - x^* \|)$.
Ceci peut s'exprimer comme suit:

$$\forall \varepsilon > 0, \exists \delta_2 > 0 \text{ tel que}$$
$$\| x - x^* \| < \delta_2 \Rightarrow \| F(x) - F(x^*) - F'(x^*; x - x^*) \| \le \varepsilon \| x - x^* \|. \quad (2.11)$$

D'autre part, le caractère semi-lisse de F en x^* implique aussi que (voir théorème 0.2, (iv)) pour tout x tendant vers x^*, pour tout V dans $\delta F(x)$,
$V(x - x^*) - F'(x^*; x - x^*) = o(\| x - x^* \|)$, c'est-à-dire

$$\forall \varepsilon > 0, \exists \delta_3 > 0 \text{ tel que} \quad (2.12)$$
$$\| x - x^* \| < \delta_3, V \in \delta F(x) \Rightarrow \| V(x - x^*) - F'(x^*; x - x^*) \| \le \varepsilon \| x - x^* \|.$$

Puisque (2.12) est vérifiée pour tout V dans $\delta F(x)$, elle l'est évidemment pour tout V appartenant à $\delta_B F(x)$.

Soit $\delta = \min \{\delta_1, \delta_2, \delta_3\}$.
En choisissant $\varepsilon = \frac{1}{3c}$ dans (2.11) et (2.12), nous pouvons maintenant prouver (2.10) par récurrence:

• pour k = 0, choisir x^o tel que $\| x^o - x^* \| < \delta$. L'égalité en (2.1) est alors bien définie dans ce cas, puisque $\delta \le \delta_1$.

• supposons que $\| x^k - x^* \| < \delta$; montrons que $\| x^{k+1} - x^* \| < \delta$.
Comme $F(x^*) = 0$, nous obtenons successivement

$$\| x^{k+1} - x^* \| = \| x^k - x^* - V_k^{-1} F(x^k) \|,$$

$$= \| - V_k^{-1} \{F(x^k) - F(x^*) - F'(x^*; x^k - x^*)\}$$

$$+ V_k^{-1} \{V_k (x^k - x^*) - F'(x^*; x^k - x^*)\} \|$$

$$\leq \| V_k^{-1} \| \left[\| F(x^k) - F(x^*) - F'(x^*; x^k - x^*) \|\right.$$

$$\left. + \| V_k (x^k - x^*) - F'(x^*; x^k - x^*) \|\right], \qquad (2.13)$$

avec $V_k \in \delta_B F(x^k)$.

L'hypothèse de récurrence ($\| x^k - x^* \| < \delta$) nous permet alors d'utiliser (2.9), (2.11) et (2.12) afin de majorer l'expression (2.13). Nous obtenons ainsi:

$$\| x^{k+1} - x^* \| \leq (2.13)$$

$$\leq c \left[\frac{1}{3c} \| x^k - x^* \| + \frac{1}{3c} \| x^k - x^* \|\right] = \frac{2}{3} \| x^k - x^* \|$$

$$< \frac{2}{3} \delta < \delta.$$

La preuve de (2.10) est donc achevée.

Nous avons en outre obtenu la convergence locale de la suite $\{x^k\}$ vers x^*. En effet, nous avons vu ci-dessus que

$$\| x^{k+1} - x^* \| \leq \frac{2}{3} \| x^k - x^* \| \quad \text{pour tout k.}$$

Par conséquent, il existe un scalaire γ, avec $0 < \gamma < 1$, tel que

$$\| x^{k+1} - x^* \| < \gamma \| x^k - x^* \| \quad \text{pour tout k.}$$

Ensuite, prouvons que le taux de convergence est bien superlinéaire.
Nous savons que pour toute suite $\{x^k\}$ convergeant vers x^* et pour tout V_k dans $\delta F(x^k)$ (donc en particulier dans $\delta_B F(x^k)$), le caractère semi-lisse de F implique que

$$F(x^k) - F(x^*) - F'(x^*; x^k - x^*) = o (\| x^k - x^* \|)$$

et $\qquad V_k(x^k - x^*) - F'(x^*; x^k - x^*) = o (\| x^k - x^* \|).$

Donc,

$$\| x^{k+1} - x^* \| \leq (2.13)$$

$$\leq c \, [\, o(\| x^k - x^* \|) + o(\| x^k - x^* \|) \,]$$

$$= o(\| x^k - x^* \|).$$

Maintenant, montrons que l'égalité (2.8) est vérifiée, c'est-à-dire que

$$\lim_{k\to\infty} \frac{\| F(x^{k+1}) \|}{\| F(x^k) \|} = 0.$$

Comme la convergence de la suite $\{x^k\}$ vers x^* est superlinéaire, nous avons que pour tout ε strictement positif, il existe un scalaire $\delta \leq \min\{\delta_1, \delta_2\}$ tel que

$$\| x^k - x^* \| < \delta \Rightarrow \| x^{k+1} - x^* \| \leq \varepsilon \| x^k - x^* \|. \qquad (2.14)$$

Or, à partir d'un certain indice k_o, tous les éléments de la suite des itérés $\{x^k\}$ vérifient $\| x^k - x^* \| < \delta$.
Nous obtenons donc successivement que pour tout indice k plus grand que k_o,

$$\| F(x^{k+1}) \| \leq \| F'(x^*; x^{k+1} - x^*) \| + \| F(x^{k+1}) - F'(x^*; x^{k+1} - x^*) \|$$

$$\leq \| F'(x^*; x^{k+1} - x^*) \| + \varepsilon \| x^{k+1} - x^* \|,$$

par (2.11), où $F(x^*) = 0$,

$$\leq (L + \varepsilon) \| x^{k+1} - x^* \|$$

$$\leq (L + \varepsilon)\, \varepsilon \| x^k - x^* \|, \quad \text{par (2.14)}, \qquad (2.15)$$

où L est la constante de Lipschitz de F au voisinage de x^*.

D'autre part, nous avons aussi que

$$\| x^k - x^* \| \leq \| x^{k+1} - x^k \| + \| x^{k+1} - x^* \|$$

$$= \| V_k^{-1} F(x^k) \| + \| x^{k+1} - x^* \| \qquad \text{par (2.1)}$$

$$\leq c \| F(x^k) \| + \varepsilon \| x^k - x^* \| \qquad \text{par (2.9) et (2.14)},$$

d'où

$$\| x^k - x^* \| \leq \frac{c}{1 - \varepsilon} \| F(x^k) \|. \tag{2.16}$$

Les relations (2.15) et (2.16) entraînent finalement que

$$\begin{aligned} \| F(x^{k+1}) \| &\leq (L + \varepsilon)\, \varepsilon \, \| x^k - x^* \| \\ &\leq \frac{c\,(L + \varepsilon)\, \varepsilon}{1 - \varepsilon} \| F(x^k) \|. \end{aligned} \tag{2.17}$$

Prenons la limite lorsque k tend vers l'infini dans l'inégalité ci-dessus: il en résulte que la relation (2.8) est vérifiée, puisque $F(x^k) \neq 0$ pour tout k et ε peut être arbitrairement petit.

Enfin, si $F'(\bullet; \bullet)$ est semi-continue de degré 2 en x^*, il suit de la proposition 0.5 appliquée à (2.13) que le taux de convergence est quadratique.
En effet,

$$\begin{aligned} \| x^{k+1} - x^* \| &\leq \| V_k^{-1} \| \big[\| F(x^k) - F(x^*) - F'(x^*; x^k - x^*) \| \\ &\qquad + \| V_k (x^k - x^*) - F'(x^*; x^k - x^*) \| \big] \qquad (2.18) \\ &\leq c \big[o(\| x^k - x^* \|) + O(\| x^k - x^* \|^2) \big] \\ &\leq O(\| x^k - x^* \|^2). \end{aligned}$$

Ceci achève la preuve. □

Corollaire 2.2

Soient $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et x^* une solution de l'équation $F(x) = 0$.
Supposons que F soit semi-lisse et fortement BD-régulière en x^*.
Alors, pour tout ε positif, il existe un δ strictement positif tel que pour tout x vérifiant $\| x - x^* \| \leq \delta$ et pour tout V appartenant à $\delta_B F(x)$, les assertions suivantes soient vérifiées:

(i) V est non singulière;
(ii) $\| x - V^{-1} F(x) - x^* \| \leq \varepsilon \| x - x^* \|$;
(iii) $\| F(x - V^{-1} F(x)) \| \leq \varepsilon \| F(x) \|$.

PREUVE:

La proposition 2.1 assure qu'il existe un voisinage U de x^* et une constante c tels que pour tout x dans U et pour tout V dans $\delta_B F(x)$, V est non singulière et $\| V^{-1} \| \leq c$, c'est-à-dire:

$$\exists\, \delta_1 > 0 \text{ tel que } \forall\, x, \tag{2.19}$$

$$\| x - x^* \| \leq \delta_1 \text{ et } V \in \delta_B F(x) \Rightarrow V \text{ est non singulière et } \| V^{-1} \| \leq c.$$

D'autre part, la suite des itérés $\{x^k\}$ produite par (2.1) convergeant vers x^* de manière superlinéaire, nous avons que

$$\forall\, \varepsilon > 0, \exists\, \delta_2 > 0 \text{ tel que}$$

$$\| x^k - x^* \| < \delta_2 \Rightarrow \| x^{k+1} - x^* \| \leq \varepsilon \| x^k - x^* \|,$$

ou encore:

$$\forall\, \varepsilon > 0, \exists\, \delta_2 > 0 \text{ tel que}$$

$$\| x^k - x^* \| < \delta_2 \Rightarrow \| x^k - V_k^{-1} F(x^k) - x^* \| \leq \varepsilon \| x^k - x^* \|, \tag{2.20}$$

où $V_k \in \delta_B F(x^k)$.

Ensuite, l'inégalité (2.17) assure que pour x^k assez proche de x^*,

$$\| F(x^{k+1}) \| \leq \frac{c\, \varepsilon\, (L + \varepsilon)}{1 - \varepsilon} \| F(x^k) \|,$$

où ε peut être arbitrairement petit, c'est-à-dire

$$\| F(x^k - V_k^{-1} F(x^k)) \| \leq \frac{c\, \varepsilon\, (L + \varepsilon)}{1 - \varepsilon} \| F(x^k) \|, \tag{2.21}$$

pour x^k tel que la norme $\| x^k - x^* \|$ soit strictement plus petite qu'un certain δ_3.

Il résulte de (2.19), (2.20) et (2.21) que pour tout ε positif, il existe un réel δ strictement positif tel que pour tout x satisfaisant à $\| x - x^* \| \leq \delta$ et pour tout V dans $\delta_B F(x)$, la thèse soit vérifiée. □

2.3 Méthode basée sur la B-dérivée

Le théorème 2.1 nous permet aussi d'obtenir des résultats valables pour la méthode de Newton basée sur la B-dérivée, c'est-à-dire la méthode définie par les équations suivantes:

$$x^{k+1} = x^k + d^k, \tag{1.11}$$

$$F(x^k) + F'(x^k; d^k) = 0. \tag{1.12}$$

Ces résultats sont établis par les corollaires 2.3 et 2.4.

Corollaire 2.3

> Sous les conditions du théorème 2.1, s'il est toujours possible de résoudre (1.12), alors toutes les conclusions du théorème sont valables pour le système (1.11)–(1.12).

PREUVE:
D'après les définitions de $\delta_B F(x^k)$ et du caractère semi-lisse, si l'équation (1.12) est toujours résolvable, il existe une matrice V_k dans $\delta_B F(x^k)$ telle que

$$F'(x^k; d^k) = V_k\, d^k.$$

La méthode définie par (1.11) et (1.12) revient alors à la méthode

$$x^{k+1} = x^k + d^k,$$
$$F(x^k) + V_k\, d^k = 0,$$

c'est-à-dire la méthode définie par (2.1). Il suffit alors d'appliquer le théorème 2.1 pour obtenir la conclusion. □

Corollaire 2.4

> Si x^* est un zéro de F et si F est semi-lisse et fortement BD-régulière en x^*, alors, pour tout ε positif, il existe un scalaire δ strictement positif tel que pour tout x vérifiant $\| x - x^* \| \leq \delta$, si l'équation $F(x) + F'(x; d) = 0$ est résolvable pour tout d, nous avons les résultats suivants:

$$\| x + d - x^* \| \leq \varepsilon \| x - x^* \|$$

et

$$\| F(x + d) \| \leq \varepsilon \| F(x) \|.$$

PREUVE:
Comme il existe une matrice V dans $\delta_B F(x)$ telle que $F'(x; d) = V\, d$, il suit de l'équation (1.12) que $d = - V^{-1} F(x)$. Nous obtenons ensuite la thèse grâce au corollaire 2.2:

$$\| x + d - x^* \| = \| x - V^{-1} F(x) - x^* \| \leq \varepsilon \| x - x^* \|$$

et

$$\| F(x + d) \| = \| F(x - V^{-1} F(x)) \| \leq \varepsilon \| F(x) \|. \qquad \square$$

CHAPITRE 3

Généralisation de la méthode de Newton amortie

Considérons l'équation $F(x) = 0$, où $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$ est une fonction localement lipschitzienne. L'objectif de ce chapitre est de présenter une généralisation de la méthode de Newton amortie, afin de résoudre l'équation citée à l'aide d'un algorithme globalement convergent. Un problème équivalent à la résolution de $F(x) = 0$ consiste à trouver un minimum global x^* pour la fonction

$$\theta(x) = \frac{1}{2} F(x)^T F(x),$$

vérifiant $\theta(x^*) = 0$. $\theta: \mathbf{R}^n \rightarrow \mathbf{R}$ est appelée fonction norme de F.

Lorsque la fonction F est différentiable au sens de Fréchet, il suffit d'utiliser la méthode de Newton amortie classique, rappelée dans la première section de ce chapitre.

Si la fonction donnée est non lisse, la matrice jacobienne n'est plus toujours définie et la méthode classique n'est plus applicable. Nous verrons alors dans la seconde section un moyen de remédier à cet inconvénient, en supposant que F soit différentiable dans toutes les directions. Un algorithme sera proposé, ainsi qu'une étude de sa convergence.

Ensuite, une troisième section présentera une généralisation de l'algorithme précédent, permettant de ne pas tenir compte de la dérivée directionnelle de la fonction F. Cette dérivée peut en effet représenter une limitation non négligeable pour atteindre la convergence globale.

Nous aborderons enfin les propriétés de convergence globale de l'algorithme généralisé, dans une dernière section.

3.1 La méthode de Newton amortie

Nous rappelons ici la méthode de Newton amortie classique, utilisée en vue de résoudre le système d'équations $F(x) = 0$ pour une fonction F localement lipschitzienne, lorsque cette fonction est différentiable au sens de Fréchet. Dans ce cas, chaque itération consiste à résoudre l'équation de Newton

$$F(x^k) + JF(x^k)\, d^k = 0 \tag{3.1}$$

pour obtenir une direction d^k et ensuite à poser

$$x^{k+1} = x^k + \tau_k\, d^k,$$

où τ_k appartenant à l'intervalle semi-ouvert (0, 1] est la longueur du pas, choisie de telle sorte que la fonction norme θ vérifie $\theta(x^{k+1}) < \theta(x^k)$.

3.2 La méthode de Newton amortie, appliquée au cas non lisse

La difficulté majeure rencontrée lors de l'application de cette méthode au cas non lisse réside dans la résolution de l'équation de Newton (3.1) afin d'obtenir la direction de recherche. En effet, la matrice jacobienne $JF(x^k)$ n'est pas toujours définie pour une fonction F non lisse.

Nous avons vu dans les chapitres précédents qu'une manière de remédier à la difficulté évoquée ci-dessus était de remplacer l'équation de Newton (3.1) par l'égalité

$$F(x^k) + V_k\, d^k = 0, \tag{3.2}$$

avec V_k dans $\delta F(x^k)$.

Mais, outre le fait que cette méthode n'ait pas de bonnes propriétés de convergence globale, un inconvénient potentiel de cette approche est la dépendance envers le jocobien généralisé, qui peut être difficile à calculer.

Pour ces raisons, nous travaillerons avec une méthode dont nous savons qu'elle peut être considérée comme un cas spécial de (3.2): la méthode basée sur la B-dérivée, où l'on a remplacé l'équation (3.1) par la relation

$$F(x^k) + F'(x^k; d^k) = 0, \qquad (3.3)$$

en supposant que la fonction F soit différentiable dans toutes les directions, en plus d'être localement lipschitzienne (voir chapitre 1).

Nous obtenons alors un algorithme applicable au cas non lisse et permettant d'atteindre une convergence globale.

<u>Algorithme 3.1</u> [Méthode de Newton amortie, cas non lisse]

<u>Pas 1</u>: Soient un vecteur arbitraire x^o et des scalaires s, β et σ donnés tels que $s > 0$, $\beta \in (0, 1)$ et $\sigma \in (0, 1/2)$.

<u>Pas 2</u>: Etant donné x^k vérifiant $F(x^k) \neq 0$, résoudre l'équation de Newton généralisée

$$F(x^k) + F'(x^k; d^k) = 0$$

afin d'obtenir une direction d^k.
Calculer la longueur du pas $\tau_k = \beta^{m_k} s$, où m_k est le plus petit entier non négatif m pour lequel

$$\theta(x^k) - \theta(x^k + \beta^m s\, d^k) \geq - \sigma\, \beta^m s\, \theta'(x^k; d^k). \qquad (3.4)$$

<u>Pas 3</u>: Poser $x^{k+1} = x^k + \tau_k d^k$ et retourner au pas 2.

Un théorème de convergence globale a été donné par Pang pour cette méthode: nous en reprenons l'énoncé ci-dessous (voir [5] pour plus de détails).

Théorème 3.1 **[Convergence globale, Pang]**

Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne.

Supposons que F soit B-différentiable. Imposons en outre les conditions suivantes:

(a) l'ensemble de niveau $\{x \in \mathbf{R}^n : \| F(x) \| \leq \| F(x^0) \| \}$ est borné;

(b) pour tout k dans $\mathbf{N}$, l'équation (3.3) a une solution.

Soit $\{x^k\}$ une suite générée par l'algorithme 3.1. Supposons que pour tout k, $F(x^k) \neq 0$.

Nous obtenons alors les conclusions suivantes:

(i) $\forall\, k \in \mathbf{N},\ \| F(x^{k+1}) \| < \| F(x^k) \|$;

(ii) la suite $\{x^k\}$ est bornée;

(iii) si $\liminf_{k\to\infty} \tau_k > 0$, alors tout point d'accumulation de la suite $\{x^k\}$ est un zéro de F.

Supposons maintenant que x^* soit un point d'accumulation de $\{x^k\}$. Si

(c) la dérivée forte de Fréchet de la fonction norme θ existe en x^*,

(d) il existe un voisinage U de x^* et un scalaire c tels que pour tout vecteur x de U, pour tout h dans $\mathbf{R}^n$, $\| h \| \leq c \| F'(x; h) \|$,

alors x^* est un zéro de la fonction F.

□

La condition (c) du théorème impose l'existence de la dérivée forte de Fréchet de θ en x^*. Si x^* est une solution du système $F(x) = 0$, alors la proposition 2 des annexes assure que cette dérivée existe et que $\nabla\theta(x^*) = 0$. Cependant, on ne peut conclure cela avant de savoir que x^* est un zéro de F. Outre cette dépendance envers la dérivée de Fréchet (déjà rencontrée au chapitre 1, section 3), le théorème exige que l'équation (3.3) ait toujours une solution. Une limitation de cette méthode réside dès lors dans la discontinuité potentielle de la dérivée directionnelle de F.

Le théorème 3.1 soulève trois autres questions. Quand la suite des itérés converge-t-elle? Quel est le taux de convergence de cette suite? Enfin, sous quelle(s) condition(s) la longueur du pas devient-elle égale à 1?

Nous allons répondre à ces trois questions, avant de généraliser la méthode de telle manière que la dérivée directionnelle des fonctions concernées n'intervienne plus.

Théorème 3.2

> Dans l'algorithme 3.1, posons $s = 1$.
> Supposons que la condition (b) du théorème 3.1 soit satisfaite. Soient $\{x^k\}$ une suite générée par l'algorithme, avec $F(x^k) \neq 0$ pour tout k, et x^* un point d'accumulation de cette suite, tel que $F(x^*) = 0$.
>
> Si la fonction F est semi-lisse et fortement BD-régulière en x^*, alors toute la suite $\{x^k\}$ converge vers x^* de manière superlinéaire et la longueur du pas, τ_k, devient finalement égale à 1.
>
> En outre, si la dérivée directionnelle classique $F'(\bullet; \bullet)$ est semi-continue de degré 2 en x^*, le taux de convergence est quadratique.

PREUVE:

Dans un premier temps, posons $\overline{x}^k = x^k + d^k$, avec d^k une solution du système d'équations $F(x^k) + F'(x^k; d^k) = 0$.

D'après le corollaire 2.4, il existe un scalaire strictement positif $\overline{\delta}$ tel que pour tout x^k vérifiant $\| x^k - x^* \| \leq \delta \leq \overline{\delta}$,

$$\| \overline{x}^k - x^* \| \leq \frac{1}{2} \| x^k - x^* \| \tag{3.5}$$

et

$$\| F(\overline{x}^k) \| \leq \sqrt{1 - 2\sigma}\, \| F(x^k) \| . \tag{3.6}$$

Par conséquent, nous obtenons que

$$\theta(x^k + d^k) = \frac{1}{2} \| F(\overline{x}^k) \|^2 \leq \frac{1 - 2\sigma}{2} \| F(x^k) \|^2$$

$$= (1 - 2\sigma)\, \theta(x^k),$$

c'est-à-dire

$$\theta(x^k) - \theta(x^k + d^k) \geq 2\sigma\, \theta(x^k). \tag{3.7}$$

La proposition 3 des annexes affirme que

$$2\, \theta(x^k) = - \theta'(x^k; d^k)$$

et l'inégalité (3.7) devient donc

$$\theta(x^k) - \theta(x^k + d^k) \geq - \sigma\, \theta'(x^k; d^k).$$

La relation (3.4) entraîne alors que

$$\tau_k = 1 \tag{3.8}$$

et

$$x^{k+1} = x^k + d^k = \bar{x}^k. \tag{3.9}$$

Par (3.5), nous avons aussi que le vecteur x^{k+1} vérifie

$$\| x^{k+1} - x^* \| \leq \frac{1}{2} \| x^k - x^* \| \leq \delta. \tag{3.10}$$

Montrons maintenant que toute la suite $\{x^k\}$ converge vers x^* de manière superlinéaire. A cette fin, remarquons que puisque x^* est un point d'accumulation de la suite, il existe un indice $k(\delta)$ dépendant de δ tel que $\| x^{k(\delta)} - x^* \| \leq \delta$. Par induction des arguments ci-dessus, il suit que les relations (3.8) à (3.10) sont vérifiées par x^k pour tout indice $k \geq k(\delta)$.

Donc, pour tout scalaire strictement positif $\delta \leq \bar{\delta}$, nous avons prouvé qu'il existait un indice $k(\delta)$ assurant (3.8) à (3.10) pour tout $k \geq k(\delta)$. Nous en déduisons que toute la suite $\{x^k\}$ converge vers x^* et que la longeur du pas, τ_k, devient finalement égale à 1. Nous obtenons en outre le taux de convergence superlinéaire grâce au corollaire 2.3.

Terminerons la preuve en montrant que la convergence quadratique est établie en conséquence de la condition plus forte: cette conclusion suit à nouveau du corollaire 2.3.

□

En fait, le théorème 3.2 permet de donner un moyen de juger si un point limite x^* de la suite $\{x^k\}$ est ou non un zéro de la fonction F, sans faire référence à la différentiabilité au sens de Fréchet de F en ce point x^*. Cette affirmation est exprimée par le corollaire 3.3.

Corollaire 3.3

Dans l'algorithme 3.1, posons $s = 1$.
Supposons que la condition (b) du théorème 3.1 soit satisfaite.
Soient $\{x^k\}$ une suite générée par l'algorithme, avec $F(x^k) \neq 0$ pour tout k, et x^* un point d'accumulation de cette suite.
Supposons en plus que la fonction F soit semi-lisse et fortement BD-régulière en x^*.

Alors x^* est solution de $F(x) = 0$ si et seulement si la suite $\{x^k\}$ converge vers x^* et la longueur du pas, τ_k, devient finalement égale à 1.

D'autre part, x^* n'est pas solution de $F(x) = 0$ si et seulement si soit la suite $\{x^k\}$ diverge, soit $\lim_{k\to\infty} \tau_k = 0$.

PREUVE:
Il suit du théorème 3.2 que si la suite $\{x^k\}$ diverge, alors x^* ne peut être un zéro de la fonction F. Supposons alors que cette suite converge. Puisque x^* en est un point d'accumulation, nous obtenons que la suite converge vers x^*.

Si x^* est une solution de l'équation $F(x) = 0$, le théorème 3.2 assure que la longueur du pas τ_k devient finalement 1 et que le taux de convergence de la suite vers x^* est superlinéaire. Ceci signifie aussi que dans le cas où $\lim_{k\to\infty} \tau_k = 0$, la longueur du pas ne peut devenir 1, d'où x^* n'est pas solution de $F(x) = 0$.

D'autre part, si x^* n'est pas solution de l'équation $F(x) = 0$, nous obtenons que

$$\lim_{k\to\infty} \| F(x^k) \| = \| F(x^*) \| > 0.$$

Or, nous avons que

$$\lim_{k\to\infty} \tau_k \| F(x^k) \| = 0. \qquad \text{(voir Pang, [5], preuve du théorème 4)}$$

Par conséquent, $\lim_{k\to\infty} \tau_k = 0$.

Nous voyons aussi par là que si la suite $\{x^k\}$ converge vers x^* et si la longueur du pas devient finalement 1, x^* est une solution de $F(x) = 0$. En effet, si x^* était limite de $\{x^k\}$

mais n'était pas un zéro de F, nous venons de montrer que la suite $\{\tau_k\}$ tendrait vers zéro, ce qui contredirait le fait que la longueur du pas devient finalement 1.

La preuve des deux équivalences est ainsi terminée. □

3.3 Généralisation de la méthode de Newton amortie

La limitation de la méthode de Newton amortie basée sur l'équation (3.3) est due à la discontinuité possible de la dérivée directionnelle $F'(x; d)$ en la variable x. Afin de remédier à cet inconvénient, nous allons développer une méthode dont l'itération ne dépend pas de la dérivée directionnelle de la fonction F.

Avant cela, nous définissons quelques notions utiles.

Définition 3.1

La dérivée directionnelle supérieure de Dini d'une fonction θ: $\mathbf{R}^n \to \mathbf{R}$, en un vecteur x et le long d'une direction v appartenant à $\mathbf{R}^n$, est donnée par

$$\theta^D(x; v) = \limsup_{\lambda \downarrow 0} \frac{\theta(x + \lambda v) - \theta(x)}{\lambda} .$$

Puisque F est une fonction localement lipschitzienne, il est évident que la fonction norme θ l'est aussi. Par conséquent, la dérivée directionnelle supérieure de Dini de θ existe en tout vecteur x de $\mathbf{R}^n$ et le long de toute direction v. De plus, le caractère localement lipschitzien de θ entraîne que pour tout vecteur x appartenant à $\mathbf{R}^n$, la dérivée $\theta^D(x; v)$ est une fonction lipschitzienne en v.

En effet, soient v et w deux directions de $\mathbf{R}^n$ et L la constante de Lipschitz au voisinage de x. Nous obtenons successivement

$$\| \theta^D(x; v) - \theta^D(x; w) \|$$

$$= \| \limsup_{\lambda \downarrow 0} \left\{ \frac{\theta(x + \lambda v) - \theta(x)}{\lambda} - \frac{\theta(x + \lambda w) - \theta(x)}{\lambda} \right\} \|$$

$$= \| \limsup_{\lambda \downarrow 0} \frac{\theta(x + \lambda v) - \theta(x + \lambda w)}{\lambda} \|$$

$$\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda} \; \| \theta(x + \lambda v) - \theta(x + \lambda w) \|$$

$$\leq \limsup_{\lambda \downarrow 0} \frac{1}{\lambda} \, L \; \| \lambda v - \lambda w \|$$

$$= L \, \| v - w \| ,$$

ce qui exprime bien que la dérivée $\theta^D(x; v)$ est une fonction lipschitzienne en v.

Notons que si chaque composante F_i: $\mathbf{R}^n \to \mathbf{R}$ de F est différentiable dans toutes les directions en x, alors la fonction norme θ l'est aussi et nous avons les égalités suivantes (voir proposition 3 des annexes):

$$\theta^D(x; v) \; = \; \theta'(x; v) \; = \; F(x)^T \, F'(x; v) \; = \sum_{i\,=\,1}^{n} F_i(x) \; F_i'(x; v).$$

Considérons maintenant l'équation F(x) = 0, dans le cas où F: $\mathbf{R}^n \to \mathbf{R}^n$ est une fonction localement lipschitzienne mais pas nécessairement différentiable dans toutes les directions. Soit x^k un itéré donné. La méthode proposée est définie par les équations suivantes:

$$x^{k+1} \; = \; x^k + \tau_k \, d^k, \tag{3.11}$$

$$F(x^k) + G(x^k, d^k) \; = \; 0, \tag{3.12}$$

où G: $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ est une fonction donnée satisfaisant aux hypothèses

(H1): $\forall \, x \in \mathbf{R}^n$, $G(x,\bullet)$ est surjective;

(H2): $F(x)^T \, G(x, v) \; \geq \; \theta^D(x; v), \;\; \forall \, (x, v) \in \mathbf{R}^n \times \mathbf{R}^n$.

La première hypothèse (H1) garantit que l'équation (3.12) est résolvable, tandis que la seconde assurera que toute solution de cette équation produira une direction de descente pour la fonction norme θ en l'itéré x^k. Plus précisément, nous pouvons établir le résultat suivant.

Proposition 3.1

> Soient F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et σ un scalaire strictement compris entre 0 et 1, arbitraires.
> Supposons que x^k vérifie $F(x^k) \neq 0$ et que les hypothèses (H1) et (H2) soient satisfaites.
> Soit d^k une solution arbitraire de l'équation (3.12).
> Alors il existe un scalaire $\overline{\tau}$ strictement positif tel que pour tout τ dans $[0, \overline{\tau}]$,
>
> $$\theta(x^k + \tau\, d^k) - \theta(x^k) \leq -2\,\sigma\,\tau\,\theta(x^k).$$

PREUVE:
Supposons par l'absurde que la conclusion soit fausse. Dans ce cas, il existe une suite $\{\tau_j\}$ de scalaires positifs, convergeant vers zéro, telle que nous ayons

$$\theta(x^k + \tau_j\, d^k) - \theta(x^k) > -2\,\sigma\,\tau_j\,\theta(x^k),$$

pour tout $j \in \mathbf{N}$.
Divisons les deux membres de l'inégalité par τ_j et passons à la limite lorsque j tend vers l'infini. Il suit de la définition de la dérivée supérieure de Dini, $\theta^D(x^k; d^k)$, que

$$\theta^D(x^k; d^k) \geq -2\,\sigma\,\theta(x^k). \tag{3.13}$$

D'autre part, l'équation (3.12) et la condition (H2) entraînent que

$$\begin{aligned}\theta^D(x^k; d^k) &\leq F(x^k)^T G(x^k, d^k)\\ &= -F(x^k)^T F(x^k)\\ &= -2\,\theta(x^k).\end{aligned} \tag{3.14}$$

En comparant les inégalités (3.13) et (3.14), nous obtenons une contradiction avec le choix de σ et l'hypothèse selon laquelle $F(x^k) \neq 0$.
Ceci termine la preuve. □

La proposition 3.1 nous permet d'établir un algorithme pour résoudre le problème de départ, à savoir résoudre le système d'équations $F(x) = 0$, où F une est fonction localement lipschitzienne.

Algorithme 3.2 [Méthode de Newton amortie, généralisée]

Pas 1: Soient un vecteur $x^o \in \mathbf{R}^n$ et deux scalaires donnés ρ et σ dans l'intervalle ouvert (0, 1).

Pas 2: Etant donné x^k vérifiant $F(x^k) \neq 0$, résoudre l'équation

$$F(x^k) + G(x^k, d^k) = 0$$

afin d'obtenir une direction d^k.

Calculer la longueur du pas $\tau_k = \rho^{m_k}$, où m_k est le plus petit entier non négatif m pour lequel

$$\theta(x^k + \rho^m d^k) - \theta(x^k) \leq -2\,\sigma\,\rho^m\,\theta(x^k). \tag{3.15}$$

Pas 3: Poser $x^{k+1} = x^k + \tau_k d^k$, tester x^{k+1} pour la convergence et retourner au pas 2.

Remarquons que dans l'algorithme, la recherche linéaire de la longueur du pas suit de la règle d'Armijo; d'après la proposition 3.1, l'entier m_k peut être déterminé après un nombre fini d'essais, en commençant avec m = 0.

3.4 Convergence de la méthode généralisée

Sous les hypothèses (H1) et (H2), la métode de Newton généralisée génère une suite de vecteurs $\{x^k\}$ qui vérifient

$$\theta(x^{k+1}) \leq (1 - 2\sigma\,\tau_k)\,\theta(x^k) < \theta(x^k),$$

où $\tau_k = \rho^{m_k}$ est la longueur du pas à la $k^{ème}$ itération.

Sans perdre de généralité, nous supposons que $\theta(x^k)$ soit strictement positive pour tout k (sinon, la norme vaudrait zéro et le problème serait résolu). Alors, la suite $\{\theta(x^k)\}$ est strictement décroissante et bornée inférieurement par zéro. Par conséquent, cette suite converge et la suite $\{\theta(x^{k+1}) - \theta(x^k)\}$ tend vers zéro.

La règle permettant de calculer la longueur du pas, (3.15), implique donc que

$$\lim_{k\to\infty} \tau_k\, \theta(x^k) = 0.$$

En effet, il suffit pour voir cela de passer à la limite dans les inégalités suivantes:

$$\frac{-1}{2\sigma}\,(\theta(x^{k+1}) - \theta(x^k)) \geq \tau_k\, \theta(x^k) \geq 0.$$

D'où, si

$$\limsup_{k\to\infty} \tau_k > 0,$$

alors il suit que $\lim_{k\to\infty} \theta(x^k) = 0$, puisque la suite $\{\theta(x^k)\}$ converge. Ceci entraîne que si un certaine sous-suite de pas $\{\tau_k\}$ est bornée, contenue dans un intervalle ne comprenant pas zéro, alors la suite entière $\{\theta(x^k)\}$ doit tendre vers zéro. En particulier, si la suite $\{x^k\}$ est aussi bornée, alors chacun de ses points d'accumulation est un zéro de la fonction F.

Afin de compléter l'analyse de convergence, nous supposerons dorénavant que la suite des itérés $\{x^k\}$ soit bornée. Il est intéressant de remarquer que cette condition est satisfaite si l'ensemble de niveau

$$\{x \in \mathbf{R}^n : \theta(x) \leq \theta(x^o) \}$$

est borné.
Cette dernière condition est vérifiée si la fonction F est coercive, c'est-à-dire si

$$\lim_{\|x\|\to\infty} \| F(x) \| = \infty.$$

En effet, supposons par l'absurde que l'ensemble de niveau $F_\alpha \equiv \{x \in \mathbf{R}^n : \theta(x) \leq \alpha \}$ soit non borné. Alors, il existe une suite $\{x^k\}$ contenue dans F_α telle que la suite des normes $\{\|x^k\|\}$ tende vers l'infini. Comme la fonction F est coercive, nous obtenons que $\lim_{k\to\infty} \| F(x^k) \| = \infty$, ce qui contredit le fait que la suite $\{x^k\}$ est dans l'ensemble F_α.

Soit x^* un point d'accumulation quelconque d'une suite d'itérés $\{x^k\}$. Nous allons montrer que x^* est un zéro de la fonction F si les deux conditions suivantes sont vérifiées:

(H3): il existe un voisinage U de x^* et une constante $L > 0$ tels que

$\forall z \in U, \forall v \in \mathbf{R}^n, \ \| G(z, v) \| \geq L \| v \|$;

(H4): pour toute suite $\{z^k\}$ convergeant vers x^*, toute suite $\{\lambda_k\}$ de scalaires strictement positifs tendant vers zéro et toute suite $\{v^k\}$ convergente,

$$\lim_{k\to\infty} F(z^k)^T G(z^k, v^k) \geq \limsup_{k\to\infty} \frac{\theta(z^k + \lambda_k v^k) - \theta(z^k)}{\lambda_k},$$

chaque fois que la limite du membre de gauche existe.

Considérons ces deux conditions dans le cas où F est une fonction lisse. Supposons que F ait une dérivée (de Fréchet) continue au voisinage de x^* et que $G(x, v) = JF(x)\, v$. Dans ces conditions, l'hypothèse (H3) est équivalente à la non singularité de la matrice jacobienne $JF(x^*)$; la condition limite (H4) est vérifiée en tant qu'égalité grâce à la continuité de $JF(x)$. Nous voyons par là que les hypothèses (H3) et (H4) sont des généralisations des exigences familières imposées afin d'obtenir la convergence de méthodes de Newton classiques, lors de la résolution d'équations lisses.

Notons encore que la condition (H4) est un renforcement de la condition (H2) au point x^*. Pour voir ceci, il suffit de remarquer qu'en posant $z^k = x^*$ et $v^k = v$ pour tout k, l'hypothèse (H4) se réduit à (H2) en x^*.

Le théorème suivant exprime le principal résultat de convergence concernant la méthode de Newton amortie généralisée.

<u>Théorème 3.4</u>

Soient $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et $G: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ une fonction satisfaisant aux hypothèses (H1) et (H2).
Soit $\{x^k\}$ une suite générée par l'algorithme 3.2.
Nous obtenons alors les conclusions suivantes:

(a) si $\limsup_{k\to\infty} \tau_k > 0$, alors $\lim_{k\to\infty} F(x^k) = 0$;

(b) si $\limsup_{k\to\infty} \tau_k = 0$ et si x^* est un point d'accumulation de la suite $\{x^k\}$, où les conditions (H3) et (H4) sont satisfaites, alors $F(x^*) = 0$.

PREUVE:
D'après la discussion vue plus haut, il suffit de prouver la partie (b) du théorème. Supposons donc que

$$\limsup_{k\to\infty} \tau_k = 0.$$

Ceci entraîne que

$$\lim_{k\to\infty} \tau_k = 0$$

ou, ce qui est équivalent,

$$\lim_{k\to\infty} m_k = \infty, \qquad \text{car } \tau_k = \rho^{m_k}, \text{ où } \rho \in (0, 1).$$

Puisque m_k est le plus petit entier m qui vérifie l'inégalité (3.15), il suit que pour tout k,

$$\theta(x^k + \overline{\tau}_k\, d^k) - \theta(x^k) > -2\,\sigma\,\overline{\tau}_k\, \theta(x^k), \tag{3.16}$$

où $\overline{\tau}_k = \rho^{m_k - 1}$, avec $\lim_{k\to\infty} \overline{\tau}_k = 0$, car $\overline{\tau}_k = \dfrac{\rho^{m_k}}{\rho}$ et $\rho^{m_k} \to 0$.

Soit $\{x^k : k \in K\}$ une sous-suite convergeant vers x^*. Alors, il existe un indice k_o dans K tel que pour tout $k \geq k_o$, avec $k \in K$, x^k appartient au voisinage U donné par la condition (H3). Ceci entraîne que pour de tels indices k,

$$L\, \| d^k \| \leq \| G(x^k, d^k) \| = \| F(x^k) \|.$$

Il en découle que la sous-suite de directions $\{d^k : k \in K\}$ est bornée, puisque la sous-suite $\{x^k : k \in K\}$ l'est. En passant si nécessaire à une sous-suite, nous pouvons affirmer que $\{d^k : k \in K\}$ converge vers une certaine limite d.
Divisons les deux membres de l'inégalité (3.16) par $\overline{\tau}_k$ et passons à la limite pour k tendant vers l'infini, avec k appartenant à K: nous en déduisons que

$$\limsup_{k\to\infty,\, k\in K} \frac{\theta(x^k + \overline{\tau}_k\, d^k) - \theta(x^k)}{\overline{\tau}_k} \geq -2\,\sigma\,\theta(x^*).$$

Grâce à la condition (H4) et à l'équation (3.12), nous savons que

$$\limsup_{k\to\infty,\, k\in K} \frac{\theta(x^k + \overline{\tau}_k\, d^k) - \theta(x^k)}{\overline{\tau}_k} \leq \lim_{\substack{k\to\infty \\ k\in K}} F(x^k)^T G(x^k, d^k)$$

$$= -2\,\theta(x^*).$$

Nous obtenons ainsi

$$-2\,\sigma\,\theta(x^*) \leq -2\,\theta(x^*).$$

Par conséquent, il suit de la définition de σ que $\theta(x^*) = 0$, comme souhaité.
Le théorème est ainsi prouvé. □

Nous concluons ce chapitre en constatant que les deux conditions (H3) et (H4) sont nécessaires pour la partie (b) du théorème. Lorsque la suite des longueurs de pas $\{\tau_k\}$ est bornée dans un intervalle ne comprenant pas zéro, la suite correspondante $\{\theta(x^k)\}$ converge vers zéro sans hypothèses spécifiques (voir partie (a) du théorème).

Le cas le plus délicat est celui où la longueur du pas approche zéro: c'est dans une telle situation que les conditions (H3) et (H4) sont requises.

CHAPITRE 4

Algorithme hybride

Ce chapitre a pour objectif de présenter un algorithme construit sur base de méthodes vues aux chapitres précédents, de telle sorte qu'il converge globalement et, localement, de manière quadratique.

Dans un premier temps, nous proposerons un théorème général dit "théorème d'attraction": ce théorème exprime qu'un zéro de la fonction F peut, dans un sens, attirer vers lui une suite d'itérés. Nous montrerons que ce théorème est applicable aux algorithmes 3.1 et 3.2 du chapitre précédent.

Ensuite, nous construirons un algorithme qui conservera les propriétés de convergence locale superlinéaire et quadratique de la méthode étudiée dans le chapitre 2, ainsi que la propriété de convergence globale de la méthode de Newton amortie généralisée, vue au chapitre 3. Le théorème d'attraction assurera alors la convergence d'une suite d'itérés vers une solution de l'équation $F(x) = 0$.

4.1 Un théorème d'attraction pour les cas généraux

Le théorème 3.2 du chapitre précédent affirme que la suite d'itérés $\{x^k\}$ produite par l'algorithme 3.1 (méthode de Newton amortie, cas non lisse) convergera vers un seul point d'accumulation x^*, à condition que x^* soit un zéro de la fonction F, avec F semi-lisse et fortement BD-régulière en x^*. Ceci exprime le fait que x^* est un tel zéro "fort" qu'il attire toute la suite d'itérés vers lui.

Si uniquement ce résultat d'attraction est concerné, nous pouvons laisser tomber la condition de fonction semi-lisse et réduire l'hypothèse de BD-régularité forte en BD-régularité. Ceci permet d'établir un théorème qui couvre des situations plus générales que celles abordées jusqu'à présent.

<u>Théorème 4.1</u> [Théorème d'attraction]

Soient $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et B-différentiable, et $\{x^k\}$ une suite générée par une méthode de direction de descente d'itération principale

$$x^{k+1} = x^k + \tau_k d^k,$$

de telle façon que

$$\| F(x^{k+1}) \| \leq \| F(x^k) \| . \tag{4.1}$$

Soit x^* un point d'accumulation de la suite $\{x^k\}$. Supposons que x^* soit un zéro de F et que F soit BD-régulière en x^*.
Supposons encore qu'il existe des scalaires δ_1, s et p strictement positifs tels que pour tout $k \in \mathbf{N}$,

$$\| x^k - x^* \| \leq \delta_1$$

entraîne que

$$\tau_k \leq s \quad \text{et} \quad \| d^k \| \leq p \| F(x^k) \| . \tag{4.2}$$

Alors la suite $\{x^k\}$ converge vers x^*.

<u>PREUVE</u>:

Nous allons montrer que pour tout voisinage de x^*, il existe un indice $\bar{k}$ tel que pour tout $k \geq \bar{k}$, x^k se trouve dans le voisinage considéré.

D'une part, la proposition 0.6 assure qu'il existe un scalaire c tel que pour tout vecteur h de $\mathbf{R}^n$,

$$\| h \| \leq c \| F'(x^*; h) \|. \tag{4.3}$$

D'autre part, le caractère B-différentiable de F en x^* signifie que pour ε arbitraire strictement positif, il existe un scalaire δ tel que pour tout h avec $0 < \| h \| < \delta$,

$$\| F(x^* + h) - F(x^*) - F'(x^*; h) \| < \varepsilon \| h \|.$$

Soit $\varepsilon = \frac{1}{2c}$ et posons $h = x - x^*$.
Nous obtenons alors qu'il existe un scalaire δ, pour lequel nous imposons en particulier $\delta \in (0, \delta_1)$, tel que pour tout vecteur x vérifiant $0 < \| x - x^* \| < \delta$,

$$\| F(x) - F(x^*) - F'(x^*; x - x^*) \| < \frac{1}{2c} \| x - x^* \|.$$

Par conséquent, nous avons successivement

$$\| F'(x^*; x - x^*) \| = \| F(x) - F(x) + F(x^*) + F'(x^*; x - x^*) \|,$$

où $F(x^*) = 0$, par hypothèse,

$$\leq \| F(x) \| + \| F(x) - F(x^*) - F'(x^*; x - x^*) \|$$

$$\leq \| F(x) \| + \frac{1}{2c} \| x - x^* \|. \tag{4.4}$$

Nous déduisons ensuite des inégalités (4.3) et (4.4) que

$$\| x - x^* \| \leq c \| F'(x^*; x - x^*) \|$$

$$\leq c \| F(x) \| + \frac{1}{2} \| x - x^* \|,$$

c'est-à-dire

$$\| x - x^* \| \leq 2c \| F(x^*) \|. \tag{4.5}$$

Fixons un scalaire ε arbitraire dans l'intervalle ouvert $(0, \delta)$ et définissons un voisinage

$$U(x^*, \varepsilon) = \{x \in \mathbf{R}^n : \| x - x^* \| \leq \varepsilon \text{ et } \| F(x) \| \leq \frac{\varepsilon}{2c + sp} \}.$$

Puisque x^* est un point d'accumulation de la suite $\{x^k\}$, il existe un indice $\bar{k} \in \mathbf{N}$ tel que l'itéré $x^{\bar{k}}$ appartienne à $U(x^*, \varepsilon)$. Dès lors, l'itéré suivant $x^{\bar{k}+1}$ se trouve aussi dans le voisinage $U(x^*, \varepsilon)$, car

$$\begin{aligned}
\| x^{\bar{k}+1} - x^* \| &= \| x^{\bar{k}} - x^* + \tau_k d^k \| \\
&\leq \| x^{\bar{k}} - x^* \| + \tau_k \| d^k \| \\
&\leq 2c \| F(x^{\bar{k}}) \| + sp \| F(x^{\bar{k}}) \|, && \text{par (4.2) et par (4.5),} \\
& && \text{car } \| x^{\bar{k}} - x^* \| \leq \varepsilon < \delta, \\
&= (2c + sp) \| F(x^{\bar{k}}) \| \\
&\leq \varepsilon, && \text{par définition de } U(x^*, \varepsilon),
\end{aligned}$$

et

$$\begin{aligned}
\| F(x^{\bar{k}+1}) \| &\leq \| F(x^{\bar{k}}) \| && \text{par (4.1)} \\
&\leq \frac{\varepsilon}{2c + sp}, && \text{car } x^{\bar{k}} \in U(x^*, \varepsilon).
\end{aligned}$$

Donc, nous avons que si $x^{\bar{k}}$ appartient à $U(x^*, \varepsilon)$, $x^{\bar{k}+1}$ aussi. Nous en déduisons que pour tout $k \geq \bar{k}$, le vecteur x^k se trouve dans le voisinage cité. Comme le scalaire ε est arbitraire, nous pouvons conclure que toute la suite $\{x^k\}$ converge bien vers la solution x^* de l'équation $F(x) = 0$. □

Maintenant, nous allons montrer que le théorème ci-dessus est applicable aux algorithmes 3.1 (Méthode de Newton amortie, cas non lisse) et 3.2 (Méhode de Newton amortie, généralisée).

Corollaire 4.2

> Dans l'algorithme 3.1, supposons que la condition (b) du théorème 3.1 soit vérifiée, pour une fonction F: $\mathbf{R}^n \to \mathbf{R}^n$ localement lipschitzienne et B-différentiable.
> Soit $\{x^k\}$ une suite produite par l'algorithme, avec $F(x^k) \neq 0$ pour tout k.
> Supposons que x^* soit un point d'accumulation de la suite $\{x^k\}$.
> Si x^* est un zéro de la fonction F et si la condition (d) du théorème 3.1 est satisfaite par x^*, alors la suite entière $\{x^k\}$ converge vers x^*.

PREUVE:
Montrons que les hypothèses du théorème 4.1 sont satisfaites. La thèse sera alors immédiate, en vue de ce même théorème.

Pour ce qui est des conditions $\tau_k \leq s$ et (4.1), elles sont vérifiées grâce à la règle d'Armijo utilisée dans l'algorithme (voir (3.4)).
Ensuite, si la condition (d) du théorème 3.1 est satisfaite, il existe un voisinage U de x^* et un scalaire c tels que si x^k appartient à U, alors

$$\| d^k \| \leq c \| F'(x^k; d^k) \| = c \| F(x^k) \|,$$

où l'égalité est due à l'équation (3.3), à savoir $F(x^k) + F'(x^k; d^k) = 0$.
Par conséquent, l'inégalité (4.2) est vérifiée pour $p = c$.
Il reste à voir que la fonction F est BD-régulière en x^*. Ceci est évident si l'on considère la remarque qui suit la proposition 0.6: cette propriété de F suit de la condition (d) du théorème 3.1.

En conclusion, le théorème 4.1 peut être appliqué et la thèse est alors immédiate. □

Remarque
Dans un certain sens, le corollaire 4.2 est plus fort que la première partie du théorème 3.2, puisqu'il ne suppose pas que F soit semi-lisse en x^*. Cependant, ce corollaire n'assure pas que la longueur du pas deviendra finalement égale à 1. Pour la conclusion plus forte, nous avons besoin d'évoquer le théorème 3.2.

Le théorème 4.1 est également applicable à l'algorithme 3.2., où l'on supposera F BD-régulière en x^*, avec x^* un point d'accumulation d'une suite d'itérés. Dans cet algorithme, la règle d'Armijo est encore utilisée pour la recherche linéaire. Donc, nous devons simplement regarder si la condition (4.2) est satisfaite.

La direction de descente d^k était obtenue en résolvant l'équation

$$F(x^k) + G(x^k, d^k) = 0, \tag{3.12}$$

où $G: \mathbf{R}^n \times \mathbf{R}^n \rightarrow \mathbf{R}^n$ est une fonction donnée. Soit x^* un point d'accumulation d'une suite d'itérés $\{x^k\}$. Pour identifier x^* comme étant un zéro de F, une condition utilisée était l'hypothèse (H3), c'est-à-dire

il existe un voisinage U de x^* et une constante $L > 0$ tels que

$$\forall\, x \in U \text{ et } \forall\, h \in \mathbf{R}^n,\ \| G(x, h) \| \geq L \| h \|. \tag{4.6}$$

Il découle des expressions (3.12) et (4.6) que

$$\| d^k \| \leq \frac{1}{L} \| G(x^k, d^k) \| = \frac{1}{L} \| F(x^k) \|.$$

Il suffit de poser $p = \frac{1}{L}$ pour obtenir la condition (4.2).

Rappelons que l'algorithme 3.2 a été proposé pour remédier aux inconvénients de l'algorithme 3.1. Si une fonction G peut être construite afin de satisfaire les hypothèses (H1) à (H4) (voir chapitre 3), où (H3) et (H4) sont associées à un point d'accumulation x^* d'une suite d'itérés $\{x^k\}$, alors nous avons vu que x^* est une solution de l'équation $F(x) = 0$. Dans le chapitre 5, nous construirons une telle fonction G dans le cas du problème de complémentarité non linéaire.

4.2 Un algorithme convergeant globalement et localement de manière quadratique

La méthode de Newton basée sur le jacobien généralisé, définie par

$$x^{k+1} = x^k - V_k^{-1} F(x^k) \tag{2.1}$$

avec V_k appartenant à $\delta_B F(x^k)$, a de bonnes propriétés de convergence locale: lorsque l'itéré x^k est proche d'un zéro x^* de la fonction F, l'équation (2.1) est bien définie à condition que F soit fortement BD-régulière en x^*; le taux de convergence est en outre superlinéaire sous l'hypothèse F semi-lisse en x^*. Cependant, globalement, l'itération (2.1) peut ne pas donner une direction de descente.

D'autre part, l'algorithme 3.2 présenté au chapitre 3 converge globalement. Mais le taux de convergence superlinéaire n'a pas encore été établi pour cette méthode.

Un moyen classique d'améliorer cette situation est de façonner un nouvel algorithme qui serait à la fois globalement convergent et localement convergent de manière superlinéaire. Nous proposons ci-dessous une approche de ce type en combinant l'algorithme globalement convergent 3.2, avec la méthode de Newton basée sur le jacobien généralisé, d'itération (2.1). L'objectif est d'obtenir une méthode conservant les propriétés de convergence de chacune des composantes.

Algorithme 4.1 [Une méthode hybride]

Soit $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et B-différentiable.
Supposons que $G: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ soit une fonction d'itération donnée, satisfaisant aux hypothèses (H1) et (H2) du chapitre 3.

Pas 1: Soient x^0 un vecteur arbitraire de $\mathbf{R}^n$ et γ, β et σ des scalaires donnés tels que
$\gamma, \beta \in (0, 1)$ et $\sigma \in (0, \frac{1}{2})$.

Pas 2: En général, soit x^k vérifiant $F(x^k) \neq 0$.
Si il existe une matrice V_k non singulière dans $\delta_B F(x^k)$, poser

$$h^k = - V_k^{-1} F(x^k). \tag{4.7}$$

Si

$$\| F(x^k + h^k) \| \leq \gamma \| F(x^k) \|, \tag{4.8}$$

poser $d^k = h^k$ et $\tau_k = 1$. Aller au pas 4.
Sinon, aller au pas 3.

Pas 3: Pour obtenir une direction d^k, résoudre l'équation (3.12), c'est-à-dire

$$F(x^k) + G(x^k, d^k) = 0.$$

Poser $\tau_k = \beta^{m_k}$, où m_k est le plus petit entier non négatif m pour lequel

$$\theta(x^k) - \theta(x^k + \beta^m d^k) \geq 2\,\sigma\,\beta^m\,\theta(x^k) \tag{4.9}$$

et aller au pas 4.

Pas 4: Poser $x^{k+1} = x^k + \tau_k d^k$ et retourner au pas 2.

Le théorème 4.3 établit la convergence de cet algorithme.

Théorème 4.3 [Convergence globlale]

Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et B-différentiable.
Supposons que les hypothèses (H1) et (H2) soient satisfaites par la fonction d'itération G.
Soit $\{x^k\}$ une suite produite par l'algorithme 4.1, avec $F(x^k) \neq 0$ pour tout k.

Alors, pour tout k,

$$\| F(x^{k+1}) \| < \| F(x^k) \|. \tag{4.10}$$

Si l'ensemble de niveau $\{x \in \mathbf{R}^n : \| F(x) \| \leq \| F(x^o) \| \}$ est borné, alors la suite d'itérés $\{x^k\}$ est bornée.

Supposons que x^* soit un point d'accumulation de cette suite. Si le choix (4.7) est utilisé un nombre infini de fois, alors x^* est une solution de l'équation $F(x) = 0$; sinon, si les hypothèses (H3) et (H4) sont vérifiées en x^*, alors x^* est encore un zéro de F.

Si x^* est un zéro de la fonction F et si F est fortement BD-régulière en x^*, alors la suite $\{x^k\}$ converge vers x^*.

Si F est semi-lisse et fortement BD-régulière en x^*, alors x^* est solution de l'équation $F(x) = 0$ si et seulement si la suite $\{x^k\}$ converge vers x^* et il existe un indice $\bar{k} > 0$ tel que pour tout $k \geq \bar{k}$, la direction d^k est calculée par (4.7), c'est-à-dire que les relations (3.12) et (4.9) ne sont pas utilisées pour $k \geq \bar{k}$. Dans ce cas, la convergence est superlinéaire.

Si en outre la dérivée directionnelle $F'(\bullet; \bullet)$ est semi-continue de degré 2 en x^*, alors la convergence est quadratique.

PREUVE:

Montrons dans un premier temps que l'inégalité (4.10) est vérifiée.
Si la relation (4.8) est satisfaite, alors nous avons que

$$\| F(x^{k+1}) \| = \| F(x^k + h^k) \| \leq \gamma \| F(x^k) \| \qquad \text{par (4.8)}$$

$$< \| F(x^k) \|, \qquad \text{car } \gamma \in (0, 1).$$

Sinon, l'algorithme se déroule avec les expressions (3.12) et (4.9). Dans ce cas, la recherche linéaire suivant la règle d'Armijo, (4.9), garantit aussi le résultat.

Supposons ensuite que l'ensemble de niveau défini dans l'énoncé soit borné. Puisque l'inégalité (4.10) est vérifiée, les éléments de la suite $\{x^k\}$ se trouvent dans cet ensemble. Cette suite est donc elle aussi bornée.

Soit x^* un point d'accumulation de $\{x^k\}$. Si l'on passe un nombre infini de fois par le choix (4.7), l'inégalité (4.8) est vérifiée un nombre infini de fois. Ceci assure que

$$F(x^*) = \lim_{k\to\infty} F(x^k) = 0.$$

Par conséquent, x^* est bien un zéro de la fonction F.
Sinon, l'algorithme se déroule avec les relations (3.12) et (4.9). Si les hypothèses (H3) et (H4) sont satisfaites en x^*, nous savons grâce au théorème 3.4 que x^* est encore solution de l'équation $F(x) = 0$.

Supposons maintenant que x^* soit un zéro de F et que F soit fortement BD-régulière en x^*. D'une part, si la condition (4.8) est satisfaite au voisinage de x^*, il découle de l'inégalité (4.7) et de la proposition 2.1 que la direction d^k vérifie

$$\| d^k \| = \| V_k^{-1} F(x^k) \| \qquad \text{par (4.7)}$$

$$\leq \| V_k^{-1} \| \| F(x^k) \|$$

$$\leq c \| F(x^k) \|, \qquad \text{d'après la proposition 2.1;}$$

de plus, la longueur du pas τ_k étant telle que $\tau_k \leq 1$, il existe bien un scalaire s strictement positif tel que $\tau_k \leq s$. Donc, la condition (4.2) est satisfaite dans ce cas.

D'autre part, si les relations (3.12) et (4.9) sont utilisées, nous avons vu dans la section précédente que la condition (4.2) était également satisfaite.
Le théorème 4.1 assure alors que la suite $\{x^k\}$ converge vers x^*.

Si F est semi-lisse et fortement BD-régulière en x^*, les conclusions suivent immédiatement du corollaire 2.2 et du théorème 2.1.
En effet, supposons d'abord que x^* soit une solution de l'équation $F(x) = 0$. Alors, nous avons vu ci-dessus si F est fortement BD-régulière en x^*, la suite d'itérés converge vers x^*. D'où, x^* est évidemment un point d'accumulation cette suite; nous pouvons donc appliquer le théorème 2.1 et le corollaire 2.2 pour obtenir les résultats désirés.
Supposons réciproquement que la suite $\{x^k\}$ converge vers x^* et que les relations (3.12) et (4.9) ne soient plus utilisées à partir d'un certain indice. Alors, l'itération (4.7) est employée un nombre infini de fois; nous avons prouvé plus haut que dans ce cas, x^* est un zéro de la fonction F.

La preuve est ainsi terminée. □

En conclusion, le théorème 4.3 exprime le fait que l'algorithme hybride conserve les propriétés de méthodes qui le composent. En outre, ce théorème affirme que sous les conditions F semi-lisse et fortement régulière en une solution x^*, l'algorithme démarrant en un vecteur initial quelconque utilisera uniquement l'itération (4.7) à partir d'un certain rang: la convergence superlinéaire est donc assurée lorsque les itérés deviennent assez proches de la solution x^*.

CHAPITRE 5

Le problème de complémentarité non linéaire

Dans l'algorithme hybride proposé au chapitre précédent afin de résoudre l'équation $F(x) = 0$, avec $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne, il reste à préciser comment déterminer un élément de l'ensemble $\delta_B F(x)$, ainsi que comment construire une fonction d'itération $G: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$. Nous allons répondre à ces questions, pour le cas particulier du problème de complémentarité non linéaire. Ce problème consiste à trouver un vecteur x de $\mathbf{R}^n$ tel que

$$h(x) \geq 0, \qquad f(x) \geq 0, \qquad h(x)^T f(x) = 0, \tag{5.1}$$

où h et f sont deux applications continûment différentiables de $\mathbf{R}^n$ dans $\mathbf{R}^n$. Nous pouvons reformuler (5.1) en un problème équivalent, afin d'obtenir la forme étudiée jusqu'à présent: la résolution de l'équation $F(x) = 0$, avec $F: \mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne. Il suffit pour cela de poser

$$F(x) = \min \{ h(x), f(x) \}, \tag{5.2}$$

où l'opérateur "min" représente le minimum composantes par composantes de deux vecteurs.

Dans un premier temps, nous proposerons une manière de construire une matrice V appartenant à l'ensemble $\delta_B F(x)$, dans le cas du problème (5.2). Ensuite, nous définirons une fonction d'itération $G: \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$, pour le même problème.

5.1 Construction d'une matrice V appartenant à l'ensemble $\delta_B F(x)$

Considérons le problème de complémentarité non linéaire défini par (5.1). Nous définissons alors la fonction F: $\mathbf{R}^n \to \mathbf{R}^n$ par

$$F(x) = \min \{ h(x), f(x) \}.$$

Nous noterons F_i, h_i et f_i les $i^{\text{èmes}}$ composantes des fonctions F, h et f respectivement. Notre objectif est de déterminer une matrice V dans l'ensemble $\delta_B F(x)$.

Le théorème suivant (voir [5]) donne une condition nécessaire et suffisante à la différentiabilité de la fonction F.

Théorème 5.1

Soient h, f: $\mathbf{R}^n \to \mathbf{R}^n$ des fonctions différentiables (au sens de Fréchet).
Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ la fonction définie par la relation (5.2).
Définissons l'ensemble

$$\beta(x) = \{ i : 1 \leq i \leq n, \ f_i(x) = h_i(x) \}.$$

Alors, F est différentiable en x si et seulement si, pour tout indice i de $\beta(x)$, $\nabla f_i(x) = \nabla h_i(x)$. En particulier, cette condition est vérifiée si $\beta(x)$ est vide.

□

Il suit du théorème 5.1 que la composante F_i de F est différentiable en x si et seulement si soit $h_i(x) \neq f_i(x)$, soit $h_i(x) = f_i(x)$ et $\nabla h_i(x) = \nabla f_i(x)$.

Nous pouvons maintenant proposer une procédure itérative permettant de construire une matrice V appartenant à $\delta_B F(x)$.

Considérons dans un premier temps les indices i $(1 \leq i \leq n)$ tels que la fonction F_i soit différentiable en x. Soit V_i la $i^{\text{ème}}$ colonne de la matrice V. Il suffit alors de définir V_i par

$$V_i = \nabla F_i(x).$$

En réalité,

$$
\begin{array}{ll}
\text{poser } V_i = \nabla F_i(x) = \nabla h_i(x), & \text{si } h_i(x) < f_i(x); \\
\text{poser } V_i = \nabla F_i(x) = \nabla f_i(x), & \text{si } f_i(x) < h_i(x); \\
\text{poser } V_i = \nabla F_i(x) = \nabla f_i(x) = \nabla h_i(x), & \text{si } f_i(x) = h_i(x) \\
 & \text{et } \nabla f_i(x) = \nabla h_i(x).
\end{array}
$$

Ensuite, notons I l'ensemble des indices i tels que la fonction F_i ne soit pas différentiable, c'est-à-dire

$$I = \{ i : 1 \leq i \leq n,\ f_i(x) = h_i(x),\ \nabla f_i(x) \neq \nabla h_i(x) \}.$$

<u>**Pas 1**</u>**:** Si l'ensemble I est vide, alors F est différentiable en x et il est évident que $V = JF(x)$. Dans ce cas, la matrice V est donc déterminée.
Sinon, poser $I_1 = I$.

<u>**Pas 2**</u>**:** En général, soit I_k un ensemble d'indices non vide.
Choisir un vecteur $z^k \in \mathbf{R}^n$ tel que, pour au moins un indice $i \in I_k$,

$$\nabla h_i(x)^T z^k \neq \nabla f_i(x)^T z^k.$$

Par exemple, choisir $z^k = \nabla h_i(x) - \nabla f_i(x)$.

Si $i \in I_k$ avec $\nabla h_i(x)^T z^k < \nabla f_i(x)^T z^k$, poser $V_i = \nabla h_i(x)$;
si $i \in I_k$ avec $\nabla h_i(x)^T z^k > \nabla f_i(x)^T z^k$, poser $V_i = \nabla f_i(x)$.

Poser $I_{k+1} = \{ i \in I_k : \nabla h_i(x)^T z^k = \nabla f_i(x)^T z^k \}$.
Si I_{k+1} est non vide, retourner au pas 2, avec $k = k + 1$.
Sinon, la procédure est terminée et la matrice V est entièrement déterminée.

Montrons que la matrice V construite par la procédure ci-avant est entièrement déterminée et qu'elle appartient réellement à l'ensemble $\delta_B F(x)$.

D'une part, il est évident que V est entièrement définie, car elle est construite colonne par colonne, tous les indices i compris entre 1 et n étant parcourus par la procédure.

D'autre part, pour voir que V appartient bien à $\delta_B F(x)$, définissons

$$z = \sum_k \varepsilon^{k-1} z^k,$$

où ε est un nombre positif très petit. Alors, pour tout $i \in I$, nous obtiendrons

$$\nabla h_i(x)^T z \neq \nabla f_i(x)^T z. \tag{5.3}$$

Nous pourrons ensuite construire une suite $\{y^j\}$ convergeant vers x dans la direction z, c'est-à-dire

$$y^j \to x \qquad \text{et} \qquad \frac{y^j - x}{\| y^j - x \|} \to \frac{z}{\| z \|},$$

telle que F soit différentiable en y^j pour tout indice j. En outre, la suite $\{JF(y^j)\}$ convergera vers V. Par conséquent, il découlera de la définition de $\delta_B F(x)$ que V appartient bien à cet ensemble.

Prouvons maintenant ces affirmations. Pour cela, nous précisons le déroulement de la procédure itérative vue ci-avant, tout en introduisant des notations utiles.

Soit donc $I_1 = I$, pour commencer. Prenons z^1 appartenant à $\mathbf{R}^n$ tel que pour au moins un indice j dans I_1,

$$\nabla f_j(x)^T z^1 \neq \nabla h_j(x)^T z^1.$$

Notons $J_1 = \{ j \in I_1 : \nabla f_j(x)^T z^1 \neq \nabla h_j(x)^T z^1 \}$.
Nous définissons alors, pour tout j dans J_1,

$$V_j = \nabla f_j(x) \qquad \text{si } \nabla f_j(x)^T z^1 < \nabla h_j(x)^T z^1$$

et

$$V_j = \nabla h_j(x) \qquad \text{si } \nabla f_j(x)^T z^1 > \nabla h_j(x)^T z^1.$$

En outre, nous savons que pour tout j dans $I_1 \setminus J_1$,

$$\nabla f_j(x)^T z^1 = \nabla h_j(x)^T z^1.$$

En général, supposons que nous ayons $I_s = I_{s-1} \setminus J_{s-1}$. Prenons z^s dans $\mathbf{R}^n$ tel que

$$\nabla f_j(x)^T z^s \neq \nabla h_j(x)^T z^s, \quad \forall\, j \in J_s.$$

Nous définissons à nouveau V_j, j appartenant à J_s:

$$V_j = \nabla f_j(x) \qquad \text{si } \nabla f_j(x)^T z^s < \nabla h_j(x)^T z^s$$

et

$$V_j = \nabla h_j(x) \qquad \text{si } \nabla f_j(x)^T z^s > \nabla h_j(x)^T z^s.$$

De plus,

$$\nabla f_j(x)^T z^s = \nabla h_j(x)^T z^s, \quad \forall\, j \in I_s \setminus J_s$$

et

$$\nabla f_j(x)^T z^k = \nabla h_j(x)^T z^k, \quad \forall\, j \in I_s \text{ et } \forall\, k \in \{1,.., s-1\}. \tag{5.4}$$

Remarquons que la procédure se termine avec un ensemble I_m $(m \leq n)$ et le choix de z^m dans $\mathbf{R}^n$ tel que

$$\nabla f_j(x)^T z^m \neq \nabla h_j(x)^T z^m, \quad \forall\, j \in J_m = I_m;$$

nous avons enfin que pour tout indice j dans I_m et tout k dans l'ensemble $\{1,.., m-1\}$,

$$\nabla f_j(x)^T z^k = \nabla h_j(x)^T z^k.$$

Définissons la direction z par

$$z = \sum_{1 \leq k \leq m} \varepsilon^{k-1} z^k; \tag{5.5}$$

nous allons déterminer le scalaire ε strictement positif, afin que la relation (5.3) soit satisfaite par z et pour tout indice i dans I.

A cette fin, considérons j arbitraire dans J_s, pour s fixé entre 1 et m. Sans perdre de généralité, nous supposons que

$$\nabla f_j(x)^T z^s - \nabla h_j(x)^T z^s < 0.$$

Le raisonnement est analogue lorsque l'inégalité est inverse.
Nous obtenons alors successivement:

$$
\begin{aligned}
&\nabla f_j(x)^T z - \nabla h_j(x)^T z \\
&\quad = \nabla f_j(x)^T z^1 - \nabla h_j(x)^T z^1 + \varepsilon\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^2 \\
&\qquad + .. + \varepsilon^{m-1}\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^m \\
&\quad = \varepsilon^{s-1}\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^s + .. + \varepsilon^{m-1}\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^m,
\end{aligned} \tag{5.6}
$$

où la dernière égalité est due à la relation (5.4). Montrons que l'expression (5.6) est strictement inférieure à zéro; nous aurons donc que la relation (5.3) sera vérifiée par z, pour l'indice j considéré.
Comme ε est strictement positif, il faut montrer que

$$(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^s + .. + \varepsilon^{m-s}\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^m < 0.$$

Nous savons déjà par hypothèse que

$$(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^s < 0.$$

Considérons les indices k tels que $s + 1 \leq k \leq m$ et

$$(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^k > 0;$$

notons K_s l'ensemble de ces indices. Nous pouvons alors choisir un scalaire que nous appellerons $\varepsilon_{j,s}$, strictement positif et suffisamment petit, pour que nous ayons

$$\sum_{k \in K_s} \varepsilon_{j,s}^{\,k-1}\,(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^k < -(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z^s.$$

Nous en déduisons que l'expression (5.6) est strictement négative; ainsi, pour tout réel ε vérifiant $0 < \varepsilon \leq \varepsilon_{j,s}$, si z est défini par la relation (5.5), nous obtenons

$$(\nabla f_j(x)^T - \nabla h_j(x)^T)\, z < 0.$$

Le résultat (5.3) est donc vérifié pour l'indice j considéré. Il suffit ensuite de reproduire le même raisonnement pour chaque indice s compris entre 1 et m. Nous aurons alors fixé un nombre fini de scalaires strictement positifs, notés $\varepsilon_{j,s}$, avec j appartenant à J_s, pour s variant de 1 à m. Choisissons enfin

$$\varepsilon \leq \min \{ \varepsilon_{j,s} : j \in J_s,\ 1 \leq s \leq m \}.$$

Il suit de ce choix que la relation (5.3) est vérifiée pour tout j appartenant à I, par

$$z = \sum_{1 \leq k \leq m} \varepsilon^{k-1} z^k.$$

Construisons maintenant la suite $\{y^j\}$. Pour cela, considérons un indice i quelconque dans I, de telle sorte que, sans perdre de généralité,

$$(\nabla f_i(x)^T - \nabla h_i(x)^T)\, z < 0.$$

Ceci signifie que z est une direction de descente en x pour la fonction $f_i - h_i$ ou encore:

$$\exists\ \delta_i > 0 \text{ tel que } \forall\ t \in (0, \delta_i),$$
$$f_i(x + t\,z) - h_i(x + t\,z) < f_i(x) - h_i(x),$$

avec $f_i(x) - h_i(x) = 0$, puisque i appartient à I. D'où, pour tout t dans l'intervalle ouvert $(0, \delta_i)$, nous obtenons que

$$f_i(x + t\,z) < h_i(x + t\,z),$$

c'est-à-dire que F_i est différentiable en $x + t\,z$, avec t dans $(0, \delta_i)$.
De même, pour tout i compris entre 1 et n, nous pouvons définir un réel δ_i strictement positif, de telle manière que pour tout t appartenant à $(0, \delta_i)$, la fonction F_i soit différentiable en $x + t\,z$.
Il suit alors que F est différentiable en $x + t\,z$, pour tout t dans $(0, \delta)$, où nous avons choisi $\delta \leq \min \{ \delta_i : 1 \leq i \leq n \}$.

Par conséquent, en posant

$$y^j = x + t_j\, z$$

pour tout j, avec $\{t_j\}$ une suite de scalaires strictement positifs décroissant vers zéro, nous pouvons affirmer que F est différentiable en y^j à partir d'un certain rang. Il est ensuite facile de constater que la suite $\{y^j\}$ ainsi définie converge bien vers x dans la direction z.

Pour terminer, il reste à voir que $\{JF(y^j)\}$ converge vers V, ou encore que chaque composante $JF_i(y^j)$ tend vers V_i. Pour les indices i tels que F_i est différentiable en x, le résultat est évident. Pour les autres, il faut considérer le choix de V_i et utiliser la différentiablité de F_i en y^j ainsi que le caractère continûment différentiable des fonctions f_i et h_i.

Remarque

La procédure proposée dans cette section permet de voir qu'il y a au plus 2^r éléments dans $\delta_B F(x)$, où r est le nombre d'éléments de I. Ceci est une justification supplémentaire au remplacement de $\delta F(x)$ par $\delta_B F(x)$ dans l'itération principale définissant la méthode, puisque $\delta F(x)$ contient un nombre infini d'éléments lorsque r est différent de zéro.

5.2 Définition d'une fonction d'itération G

L'algorithme hybride proposé au chapitre 4 dépend de l'existence de la fonction G(•, •) satisfaisant aux hypothèses (H1) à (H4). Nous ne savons pas comment construire une telle fonction dans la situation générale où F est localement lipschitzienne. Cependant, il est possible de définir directement une fonction d'itération pour certaines classes de problèmes, notamment la résolution de l'équation F(x) = 0, avec F B-différentiable, ou plus particulièrement le problème étudié dans ce chapitre.

Considérons le problème de complémentarité non linéaire (5.1), que nous réécrivons de la manière suivante pour simplifier les notations par la suite:

$$f_1(x) \geq 0, \quad f_2(x) \geq 0, \quad f_1(x)^T f_2(x) = 0, \tag{5.7}$$

où f_1 et f_2 sont des fonctions continûment différentiables de $\mathbf{R}^n$ dans $\mathbf{R}^n$.
La fonction F peut alors être définie composantes par composantes comme suit:

$$F_i(x) = \min \{ f_{ij}(x) : j \in \{1, 2\} \}, \tag{5.8}$$

pour $i = 1,.., n$.

Soit ε un scalaire strictement positif arbitraire; définissons l'ensemble

$$J_{i,\varepsilon}(x) = \{ j \in \{1, 2\} : f_{ij}(x) \leq F_i(x) + \varepsilon \}.$$

Notons que $J_{i,\varepsilon}(x) \supseteq J_i(x)$, où

$$J_i(x) = \{ j \in \{1, 2\} : f_{ij}(x) = F_i(x) \},$$

c'est-à-dire que $J_i(x)$ est l'ensemble des indices correspondant aux fonctions minimales pour $F_i(x)$.

Dans la suite, nous aurons besoin de la dérivée directionnelle $F'(x; d)$; le théorème 5.2 (voir [5]) assure l'existence de cette dérivée et en donne l'expression.

Théorème 5.2

Soient h, f: $\mathbf{R}^n \to \mathbf{R}^n$ des fonctions différentiables (au sens de Fréchet).
Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ la fonction définie par la relation (5.2).
Alors, F est B-différentiable partout et sa dérivée directionnelle est donnée par

$$(F'(x; d))_i = \begin{cases} \nabla f_i(x)^T d & \text{si } i \in \alpha(x), \\ \min \{ \nabla f_i(x)^T d, \nabla h_i(x)^T d \} & \text{si } i \in \beta(x), \\ \nabla h_i(x)^T d & \text{si } i \in \gamma(x), \end{cases}$$

où

$$\alpha(x) = \{ i : 1 \leq i \leq n, f_i(x) < h_i(x) \},$$
$$\beta(x) = \{ i : 1 \leq i \leq n, f_i(x) = h_i(x) \},$$
$$\gamma(x) = \{ i : 1 \leq i \leq n, f_i(x) > h_i(x) \}.$$

□

Dans les notations choisies, nous obtenons alors que

$$F_i'(x; d) = \min \{ \nabla f_{ij}(x)^T d : j \in J_i(x) \}.$$

Nous pouvons maintenant définir une fonction d'itération G: $\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n$ de la manière suivante: pour $i = 1,.., n$, soit

$$G_i(x, d) = \begin{cases} F_i'(x; d) & \text{si } F_i(x) \geq 0, \\ \min \{ \nabla f_{ij}(x)^T d : j \in J_{i,\varepsilon}(x) \} & \text{si } F_i(x) < 0. \end{cases}$$

Remarquons que pour chaque vecteur x fixé, la fonction G(x, d) est concave et linéaire par morceaux en d; donc, chaque équation

$$F(x^k) + G(x^k, d^k) = 0$$

peut être résolue par un algorithme "complémentaire pivot" (pour plus de détails, voir [2]). Cette résolution n'est pas notre but ici: il nous suffit de savoir que l'équation citée est résolvable. En effet, l'hypothèse (H1) du chapitre 3, à savoir la surjectivité de la fonction G(x, •), avait été imposée dans ce but.

Nous allons montrer que les conditions (H2) et (H4) de ce même chapitre sont satisfaites par la fonction G définie ci-dessus.

Pour la condition (H3), nous citerons simplement le résultat suivant: si G(x, •) est régulière en x*, selon la définition de régularité donnée en [2], alors cette condition (H3) est satisfaite en x*.

Montrons dans un premier temps que la condition (H2) est vérifiée, c'est-à-dire que

$$F(x)^T G(x, d) \geq \theta^D(x; d), \quad \forall (x, d) \in \mathbf{R}^n \times \mathbf{R}^n.$$

En regard de la définition de G, il est facile de voir que pour chaque composante G_i, nous avons

$$F_i(x)\, G_i(x, d) \geq F_i(x)\, F_i'(x; d). \tag{5.9}$$

Il suffit, pour obtenir ce résultat, de considérer les différentes situations possibles. L'inégalité (5.9) permet alors d'affirmer que

$$F(x)^T G(x, d) \geq F(x)^T F'(x; d) = \theta'(x; d) = \theta^D(x; d);$$

la condition (H2) est donc bien satisfaite par la fonction G.

Afin de montrer que l'hypothèse (H4) est vérifiée, prenons x^*, $\{z^k\}$, $\{v^k\}$ et $\{\lambda_k\}$, de telle manière que la suite $\{z^k\}$ converge vers x^*, la suite de scalaires $\{\lambda_k\}$ décroisse vers zéro et la suite $\{v^k\}$ soit convergente. Nous devons prouver que sous ces conditions,

$$\lim_{k\to\infty} F(z^k)^T G(z^k, v^k) \geq \limsup_{k\to\infty} \frac{\theta(z^k + \lambda_k v^k) - \theta(z^k)}{\lambda_k},$$

chaque fois que la limite du membre de gauche existe. A cette fin, nous allons analyser les composantes de F.
Pour tout indice k, nous avons que

$$\frac{\theta(z^k + \lambda_k v^k) - \theta(z^k)}{\lambda_k} = \frac{1}{2}(T_{1,k} + T_{2,k} + T_{3,k}),$$

où

$$T_{1,k} = \sum_{i:\, F_i(x^*)>0} \left[(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2\right]/\lambda_k,$$

$$T_{2,k} = \sum_{i:\, F_i(x^*)=0} \left[(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2\right]/\lambda_k,$$

$$T_{3,k} = \sum_{i:\, F_i(x^*)<0} \left[(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2\right]/\lambda_k;$$

et similairement,

$$F(z^k)^T G(z^k, v^k) = \frac{1}{2}(T'_{1,k} + T'_{2,k} + T'_{3,k}),$$

où

$$T'_{1,k} = \sum_{i:\, F_i(x^*)>0} 2\, F_i(z^k)\, G_i(z^k, v^k),$$

$$T'_{2,k} = \sum_{i:\, F_i(x^*)=0} 2\, F_i(z^k)\, G_i(z^k, v^k),$$

$$T'_{3,k} = \sum_{i:\, F_i(x^*)<0} 2\, F_i(z^k)\, G_i(z^k, v^k).$$

Considérons alors un indice i pour lequel $F_i(x^*) > 0$. Nous obtenons que pour tout indice k assez grand, les expressions $F_i(z^k + \lambda_k v^k)$ et $F_i(z^k)$ sont toutes deux strictement positives; d'où, pour le même indice k, $G_i(z^k, v^k) = F_i'(z^k, v^k)$. Soit $j \in J_i(z^k)$ tel que $\nabla f_{ij}(z^k)^T v^k = F_i'(z^k; v^k)$. Alors,

$$(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2 - 2\,\lambda_k\, F_i(z^k)\, F_i'(z^k; v^k)$$

$$= (F_i(z^k + \lambda_k v^k))^2 - (f_{ij}(z^k))^2 - 2\,\lambda_k\, f_{ij}(z^k)\, F_i'(z^k; v^k)$$

$$\leq (f_{ij}(z^k + \lambda_k v^k))^2 - (f_{ij}(z^k))^2 - 2\,\lambda_k\, f_{ij}(z^k)\, \nabla f_{ij}(z^k)^T v^k,$$

car $0 < F_i(z^k + \lambda_k v^k) \leq f_{ij}(z^k + \lambda_k v^k)$.

De cette expression, il suit que pour un tel indice i,

$$\limsup_{k\to\infty} \frac{(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2 - 2\,\lambda_k\, F_i(z^k)\, F_i'(z^k; v^k)}{\lambda_k} \leq 0,$$

ce qui entraîne que

$$\limsup_{k\to\infty} (T_{1,k} - T'_{1,k}) \leq 0.$$

D'autre part, en vue de la propriété lipschitzienne des fonctions F_i, on peut montrer que

$$\lim_{k\to\infty} T_{2,k} = \lim_{k\to\infty} T'_{2,k} = 0.$$

En effet, notons L_i la constante de Lipschitz associée à la composante F_i. Nous avons alors que

$$| T_{2,k} | = \Big| \sum_{i:\, F_i(x^*)=0} \big[(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2\big]/\lambda_k \Big|$$

$$\leq \sum_{i:\, F_i(x^*)=0} \left| \frac{(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2}{\lambda_k} \right|$$

$$= \sum_{i:\, F_i(x^*)=0} \left| \frac{(F_i(z^k + \lambda_k v^k) - F_i(z^k)) * (F_i(z^k + \lambda_k v^k) + F_i(z^k))}{\lambda_k} \right|$$

$$\leq \sum_{i:\, F_i(x^*)=0} \left[\left| \frac{F_i(z^k + \lambda_k v^k) - F_i(z^k)}{\lambda_k} \right| \, | F_i(z^k + \lambda_k v^k) + F_i(z^k) | \right]$$

$$\leq \sum_{i:\, F_i(x^*)=0} \left[\frac{L_i \lambda_k \| v^k \|}{\lambda_k} \, | F_i(z^k + \lambda_k v^k) + F_i(z^k) | \right].$$

En prenant la limite lorsque k tend vers l'infini de cette dernière expression, nous obtenons que cette limite vaut zéro, car la suite $\{v^k\}$ est bornée et par continuité de F_i, nous avons

$$\lim_{k\to\infty} \| F_i(z^k + \lambda_k v^k) + F_i(z^k) \| = 2\, F_i(x^*) = 0.$$

Finalement, il reste à prouver que

$$\limsup_{k\to\infty} (T_{3,k} - T'_{3,k}) \leq 0. \tag{5.10}$$

A cette fin, remarquons les inclusions suivantes, valables pour tout indice k suffisamment grand et pour $i = 1,.., n$:

$$J_{i,\varepsilon}(z^k) \supseteq J_i(x^*) \supseteq J_i(z^k + \lambda_k v^k).$$

Donc, si k est assez grand, nous avons que pour tout i tel que $F_i(x^*) < 0$ et tout j dans $J_i(z^k + \lambda_k v^k)$,

$$\begin{aligned} F_i(z^k) &\leq f_{ij}(z^k) && \text{par définition de } F_i, \\ &< 0, && \text{car } z^k \to x^* \text{ et } j \in J_i(x^*), \\ & && \text{d'où } f_{ij}(x^*) = F_i(x^*) < 0. \end{aligned}$$

Nous en déduisons que

$$(F_i(z^k))^2 \geq (f_{ij}(z^k))^2.$$

De plus, puisqu'un tel indice j doit appartenir à l'ensemble $J_{i,\varepsilon}(z^k)$, nous pouvons affirmer que

$$F_i(z^k)\, G_i(z^k, v^k) \geq F_i(z^k)\, \nabla f_{ij}(z^k)^T v^k.$$

Par conséquent, pour les indices i et j considérés ci-dessus, nous obtenons successivement que

$$(F_i(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2 - 2\,\lambda_k\, F_i(z^k)\, G_i(z^k, v^k)$$

$$= (f_{ij}(z^k + \lambda_k v^k))^2 - (F_i(z^k))^2 - 2\,\lambda_k\, F_i(z^k)\, G_i(z^k, v^k)$$

$$\leq (f_{ij}(z^k + \lambda_k v^k))^2 - (f_{ij}(z^k))^2 - 2\,\lambda_k\, F_i(z^k)\, \nabla f_{ij}(z^k)^T v^k, \quad (5.11)$$

car $F_i(z^k) < 0$ et par définition de $G_i(z^k, v^k)$.

Enfin, la condition limite (5.10) découle immédiatement de l'inégalité (5.11); l'hypothèse (H4) est donc bien satisfaite par la fonction d'itération G que nous avons définie.

Remarque

Grâce aux notations employées dans cette section, nous pouvons généraliser facilement la définition de la fonction d'itération G utilisée, au cas d'un problème défini comme suit: pour chaque $i = 1,.., n$, les composantes de la fonction F sont données par

$$F_i(x) = \min \{ f_{ij}(x) : j \in I_i \},$$

où I_i est un ensemble fini et f_{ij}: $\mathbf{R}^n \to \mathbf{R}$ est une fonction continûment différentiable. Notons que $F_i(x) = 0$ si et seulement si $f_{ij}(x) \geq 0$ pour tout j dans I_i et au moins une fonction vérifie $f_{ij}(x) = 0$. Donc, trouver une solution zéro de la fonction F correspond dans ce cas à la résolution d'une généralisation du problème de complémentarité non linéaire (5.1).

Conclusion

Clôturons ce chapitre en répondant à une question: quelles sont les hypothèses du théorème 4.3 qui sont vérifiées par la fonction F définie en (5.2)?

Constatons d'abord que F est localement lipschitzienne et B-différentiable. En outre, nous avons pu définir une fonction d'itération G satisfaisant aux hypothèses (H1) et (H2) du chapitre 3. Nous pouvons donc appliquer l'algorithme 4.1, afin de résoudre l'équation $F(x) = 0$.

Pour ce qui est du théorème 4.3, établissant la convergence globale de l'algorithme, déterminons les hypothèses satisfaites. D'une part, nous avons prouvé que la fonction G définie dans ce chapitre vérifiait la condition (H4). D'autre part, puisque l'opérateur "min" est un opérateur semi-lisse (voir [4]), la fonction F est semi-lisse.

CONCLUSION

Tout au long de ce mémoire, nous avons étudié des versions non lisses de la méthode de Newton classique, afin de résoudre un système d'équations non linéaires défini par une fonction localement lipschitzienne. Nous avons établi la convergence de ces méthodes: les algorithmes des chapitres 1 et 2 convergeaient localement, ceux du chapitre 3, globalement. Le regroupement de deux méthodes a permis d'obtenir un algorithme convergeant globalement et localement de manière quadratique (algorithme 4.1).

La question suivante se pose alors: puisque la méthode basée sur la B-dérivée, définie par les équations

$$x^{k+1} = x^k + d^k \tag{1.11}$$

$$F(x^k) + F'(x^k; d^k) = 0, \tag{1.12}$$

peut être considérée comme un cas spécial de celle utilisant le jacobien généralisé, dont l'itération principale est déterminée par

$$x^{k+1} = x^k - V_k^{-1} F(x^k), \tag{2.1}$$

où $V_k \in \delta_B F(x^k)$ avec $\delta_B F(x^k) = \{\lim_{\substack{x_i \to x^k \\ x_i \in D_F}} JF(x_i)\}$, pourquoi ne pas avoir construit l'algorithme hybride avec la première méthode? En effet, nous avons vu au chapitre 2 que le théorème de convergence locale (théorème 1.1) établi pour la seconde méthode était encore valable pour la méthode d'itération (1.11)–(1.12), à condition que l'équation (1.12) soit résolvable.

En réalité, chaque méthode possède ses avantages et inconvénients et il est difficile ici de déterminer la plus efficace. Plus précisément, nous avons que le calcul d'un élément de l'ensemble $\delta_B F(x)$ n'est pas toujours évident; par contre, une fois ce calcul effectué, il reste à résoudre le système (2.1), qui est linéaire, alors que l'équation (1.12) ne l'est généralement pas. D'autre part, nous savons que l'itération (2.1) est bien définie au voisinage d'une solution x^* de $F(x) = 0$, à condition que F soit semi-lisse et fortement BD-régulière en x^*. Ces hypothèses n'entraînent pas nécessairement que l'équation (1.12) admette toujours une solution lorsque les itérés deviennent assez proches de x^* (un exemple est proposé en annexes). La résolvabilité de cette équation au voisinage de x^* peut toutefois être garantie sous des hypothèses de régularité.

En résumé, la méthode définie par (1.11)–(1.12) ne présente pas de désavantage majeur par rapport à celle basée sur le jacobien généralisé, d'itération (2.1).

Pour terminer, considérons la deuxième partie de l'algorithme 4.1. Dans le chapitre 3, nous avons généralisé la méthode de Newton amortie de telle manière que la dérivée directionnelle n'intervienne plus. Pour cela, nous avons introduit une fonction d'itération G et déterminé quatre hypothèses sur cette fonction, afin d'établir la convergence globale. L'algorithme hybride dépend donc de l'existence d'une telle fonction. Malheureusement, nous ne savons la définir que pour certaines classes d'applications. Il reste donc une étude à accomplir dans cette direction. En outre, les fonctions d'itération construites ne vérifient pas toujours les quatre conditions souhaitées. Dans le cas du problème de complémentarité non linéaire, par exemple, l'hypothèse (H3) n'est pas automatiquement satisfaite: il faut donc l'imposer; une autre possibilité serait de chercher à modifier la fonction d'itération.

BIBLIOGRAPHIE

[1] Clarke, F. H. (1983), Optimization and Nonsmooth Analysis, Wiley, New York.

[2] Eaves, B. C. (1974), Solving Piecewise Linear Convex Equations, *Mathematical Programming Study*, Vol. 1, 96–119.

[3] Han, S. P., Pang, J. S. and Rangaraj, N. (1992), Globally Convergent Newton Methods for Nonsmooth Equations, *Mathematics of Operations Research*, Vol. 17, 586–606.

[4] Mifflin, R. (1977), Semismooth and Semiconvex Functions in Constrained Optimization, *SIAM J. Control and Optimization*, Vol. 15, N°6.

[5] Pang, J. S. (1990), Newton's Method for B-differentiable Equations, *Mathematics of Operations Research*, Vol. 15, 311–341.

[6] Pang, J. S. (1991), A B-differentiable Equation-Based, Globally, and Locally Quadratically Convergent Algorithm for Non Linear Programs, Complementarity and Variational Inequality Problems, *Mathematical Programming*, Vol. 51, 101–131.

[7] Qi, L. and Sun, J. (1993), A Nonsmooth Version of Newton's Method, *Mathematical Programming*, Vol. 58, 353–367.

[8] Qi, L. (1993), Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations, *Mathematics of Operations Research*, Vol. 18, 227–244.

ANNEXES

<u>Proposition 1</u> [Propriétés liées à la B-différentiabilité]

Soit $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$ une fonction localement lipschitzienne en un vecteur x.

(i) Si la fonction F est différentiable (au sens de Fréchet) en x, alors elle est B-différentiable en x et $BF(x) = JF(x)$.

Inversément, si F est B-différentiable en x et si la B-dérivée $BF(x)v$ est linéaire en v, alors F est différentiable en x.

(ii) Si F est B-différentiable en x, alors la B-dérivée est unique. De plus, $BF(x)$ est lipschitzienne avec la même constante de Lipschitz que F.

(iii) Si F est B-différentiable en x, alors F est différentiable dans toutes les directions en x et $F'(x; d) = BF(x)\, d$. La réciproque de cette affirmation est vraie également.

(iv) Les règles d'addition, de soustraction ainsi que la règle de la chaîne sont applicables aux fonctions B-différentiables.

<u>Proposition 2</u>

Soit $F: \mathbf{R}^n \rightarrow \mathbf{R}^n$ une fonction localement lipschitzienne. Si x^* est un zéro de F, alors la fonction norme de F, définie par

$$\theta(x) = \frac{1}{2} F(x)^T F(x),$$

admet une dérivée forte de Fréchet en x^* et $\nabla\theta(x^*) = 0$.

Proposition 3

Soit F: $\mathbf{R}^n \to \mathbf{R}^n$ une fonction localement lipschitzienne et B-différentiable.
Soit θ: $\mathbf{R}^n \to \mathbf{R}$ la fonction norme de F, définie par

$$\theta(x) = \frac{1}{2} F(x)^T F(x).$$

Alors, θ est B-différentiable et $B\theta(x)\, d = F(x)^T BF(x)\, d$.

En outre, si $F(x) + BF(x)\, d = 0$,
alors

$$\theta'(x; d) = - F(x)^T F(x).$$

Exemple

L'exemple suivant montre qu'une fonction B-différentiable peut être semi-lisse et fortement BD-régulière en un zéro x^*, mais telle que l'équation (1.12) ne soit pas bien définie au voisinage de x^*.
Soit G: $\mathbf{R}^2 \to \mathbf{R}^2$, qui à tout $x = (u, v)^T$ associe le vecteur $(-u, u^2 - v)^T$.
Posons $x_+ = (u_+, v_+)^T$, où $a_+ = \max\{a, o\}$ pour tout nombre réel a.

Alors, la fonction F: $\mathbf{R}^2 \to \mathbf{R}^2$ définie par

$$F(x) = G(x_+) + x - x_+$$

est une application semi-lisse telle que $x^* = (0, 0)^T$ soit un zéro de F et F soit fortement BD-régulière en x^*.
Fixons maintenant deux nombres réels $\bar{u}$ et $\bar{v}$, avec $\bar{u} > 0$ et $\bar{v} = 0$. L'équation (1.12), utilisée pour obtenir la direction $(p, q)^T$ à partir de $(\bar{u}, \bar{v})^T$, est alors donnée par

$$F\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} + F'\left(\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} ; \begin{pmatrix} p \\ q \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Il faut donc résoudre le système

$$\begin{cases} -\bar{u} - p = 0 \\ \bar{u}^2 + 2\,\bar{u}\, p - q_+ + q - q_+ = 0, \end{cases}$$

ce qui entraîne que $p = -\bar{u}$, d'où $q - 2\,q_+ = \bar{u}^2$.

Cette dernière égalité est impossible, vu le choix de $\bar{u}$ (> 0). Puisque le vecteur $(\bar{u}, \bar{v})^T$ est pris arbitrairement près de x^*, nous voyons que la méthode dont l'itération est donnée par (1.11)–(1.12) n'est pas toujours bien définie au voisinage de x^*.